

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 24, 2002, 12:10:37 ; Search time 72.239 Seconds  
(without alignments)  
3854.741 Million cell updates/sec

Title: US-09-720-086-5  
Perfect score: 4909  
Sequence: 1 MPSSGPDSSSLEREDR.....MSVPTRHLPAPKEPACV 908

Scoring table: BIOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 153338381 residues  
Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USFTO.epool/US09720086/runat.18112002.092957.29457/app.query.fasta\_1.4252  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=ini -MINMATCH=0.1 -LOOPCL=0  
-LOOEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09720086.@CGN.1.1.108.@runat.18112002.092957.29457 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAPP -LARGEBUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:\*

- 1: /cgn2\_6/prodata/1/ina/5A.COMB.seq:\*
- 2: /cgn2\_6/prodata/1/ina/5B.COMB.seq:\*
- 3: /cgn2\_6/prodata/1/ina/6A.COMB.seq:\*
- 4: /cgn2\_6/prodata/1/ina/6B.COMB.seq:\*
- 5: /cgn2\_6/prodata/1/ina/PCTUS.COMB.seq:\*
- 6: /cgn2\_6/prodata/1/ina/backfillseq.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	748	15.2	2077	4	US-09-276-531-47 Sequence 47, Appl
2	165	3.4	4190	4	US-08-924-345-1 Sequence 1, Appl
3	165	3.4	35060	3	US-08-814-095-7 Sequence 1, Appl
4	165	3.4	43280	2	US-08-804-227C-1 Sequence 1, Appl
5	162	3.3	4350	4	US-09-295-593-37 Sequence 11, Appl
6	161	3.3	12001	1	US-08-458-568A-11 Sequence 75, Appl
7	159.5	3.2	2608	4	US-09-154-750A-75 Sequence 4, Appl
8	158.5	3.2	11907	4	US-08-061-376-4 Sequence 1, Appl
9	158.5	3.2	14255	1	US-08-320-559-1 Sequence 1, Appl
10	158.5	3.2	14255	1	US-08-327-392-1 Sequence 5, Appl
11	158.5	3.2	14255	1	US-08-306-691B-55 Sequence 1, Appl
12	158.5	3.2	14255	3	US-08-545-860D-1 Sequence 1, Appl

13	158.5	3.2	14255	5	PCT-US94-04496-1 Sequence 1, Appl
14	157	3.2	4201	1	US-08-080-255-4 Sequence 4, Appl
15	157	3.2	4201	3	US-08-465-713-4 Sequence 4, Appl
16	157	3.2	4201	5	PCT-US93-05857-4 Sequence 1, Appl
17	153	3.1	11220	4	US-09-105-537-32 Sequence 32, Appl
18	153	3.1	36778	4	US-09-105-537-5 Sequence 5, Appl
19	153	3.1	38506	3	US-09-320-878-19 Sequence 19, Appl
20	152.5	3.1	8438	1	US-07-945-283-1 Sequence 1, Appl
21	151.5	3.1	4852	1	US-07-853-913-3 Sequence 3, Appl
22	151.5	3.1	6755	3	US-08-931-999-4 Sequence 4, Appl
23	150.5	3.1	5267	3	US-08-976-255-2 Sequence 2, Appl
24	150	3.1	2355	4	US-08-913-159-12 Sequence 12, Appl
25	150	3.1	3773	4	US-09-130-242-1 Sequence 1, Appl
26	149	3.0	13842	4	US-09-105-537-30 Sequence 30, Appl
27	149	3.0	36778	4	US-09-105-537-5 Sequence 5, Appl
28	149	3.0	38506	3	US-09-320-878-19 Sequence 19, Appl
29	149	3.0	43280	2	US-08-804-227C-1 Sequence 1, Appl
30	148.5	3.0	1566	2	US-08-405-175A-10 Sequence 10, Appl
31	148	3.0	9551	1	US-08-056-200-93 Sequence 93, Appl
32	148	3.0	9551	2	US-08-800-644-93 Sequence 93, Appl
33	147.5	3.0	2115	3	US-09-032-365A-12 Sequence 12, Appl
34	147.5	3.0	2116	1	US-08-701-380-1 Sequence 1, Appl
35	147	3.0	2680	4	US-09-063-035-1 Sequence 1, Appl
36	147	3.0	15872	4	US-09-105-537-1 Sequence 1, Appl
37	146	3.0	1926	4	US-09-249-585A-2 Sequence 2, Appl
38	146	3.0	2580	3	US-09-050-863-2 Sequence 2, Appl
39	146	3.0	2580	4	US-09-359-081-2 Sequence 2, Appl
40	146	3.0	5452	2	US-09-130-114-1 Sequence 1, Appl
41	146	3.0	9600	4	US-08-910-647-1 Sequence 1, Appl
42	146	3.0	9600	4	US-09-620-925-1 Sequence 1, Appl
43	146	3.0	10596	1	US-07-884-811-15 Sequence 15, Appl
44	146	3.0	10596	1	US-07-885-971-15 Sequence 15, Appl
45	146	3.0	10596	1	US-08-087-783A-15 Sequence 15, Appl

#### ALIGNMENTS

RESULT 1  
US-09-276-531-47  
Sequence 47, Application US/09276531  
Patent No. 6183968  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Yue, Henry  
APPLICANT: Reddy, Roopa  
APPLICANT: Guegler, Karl J.  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION  
NUMBER OF SEQUENCES: 134  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: word perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/276,531  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/079,677  
FILING DATE: March 27, 1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:





```
?
? TELEPHONE: (248) 539-5050
? TELEFAX: (248) 539-5055
? INFORMATION FOR SEQ ID NO: 7:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 35060 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: other nucleic acid
? DESCRIPTION: /desc = "cosmid including ACHE
? HYPOTHETICAL: NO
? ANTI-SENSE: NO
? ORGANISM: Homo sapiens
? POSITION IN GENOME:
? CHROMOSOME/SEGMENT: 7q22
? FEATURE:
? NAME/KEY: promoter
? LOCATION: 4089..22464
? OTHER INFORMATION: /function= "ACHE Promotor"
? OTHER INFORMATION: /standard_name= "ACHE Promotor"
? FEATURE:
? NAME/KEY: exon
? LOCATION: 22465..22537
? OTHER INFORMATION: /functions= "non-translated"
? OTHER INFORMATION: /gene= "ACHE"
? OTHER INFORMATION: /number= 1
? FEATURE:
? NAME/KEY: exon
? LOCATION: 24090..25177
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /function= "(translation start:
? OTHER INFORMATION: 24110)"
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /gene= "ACHE"
? OTHER INFORMATION: /number= 2
? FEATURE:
? NAME/KEY: exon
? LOCATION: 25524..26009
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /gene= "ACHE"
? OTHER INFORMATION: /number= 3
? FEATURE:
? NAME/KEY: exon
? LOCATION: 27005..27274
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /gene= "ACHE"
? OTHER INFORMATION: /number= 4
? FEATURE:
? NAME/KEY: exon
? LOCATION: 27255..28007
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /gene= "ACHE"
? OTHER INFORMATION: /number= 5
? FEATURE:
? NAME/KEY: terminator
? LOCATION: 27385..27387
? FEATURE:
? NAME/KEY: exon
? LOCATION: 28008..28129
? IDENTIFICATION METHOD: experimental
? OTHER INFORMATION: /evidence= EXPERIMENTAL
? OTHER INFORMATION: /gene= "ACHE"
? OTHER INFORMATION: /number= 6
? FEATURE:
? NAME/KEY: terminator
? LOCATION: 28129..28131
? FEATURE:
? NAME/KEY: exon
? LOCATION: 33493..33591
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 4
? FEATURE:
? NAME/KEY: exon
? LOCATION: 33297..33408
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 5
? FEATURE:
? NAME/KEY: exon
? LOCATION: 32959..33094
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 6
? FEATURE:
? NAME/KEY: exon
? LOCATION: 32569..32628
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 7
? FEATURE:
? NAME/KEY: exon
? LOCATION: 32386..32468
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 8
? FEATURE:
? NAME/KEY: exon
? LOCATION: 31894..32080
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 9
? FEATURE:
? NAME/KEY: exon
? LOCATION: 31363..31534
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 10
? FEATURE:
? NAME/KEY: exon
? LOCATION: 31131..31284
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 11
? FEATURE:
? NAME/KEY: exon
? LOCATION: 30816..31011
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 12
? FEATURE:
? NAME/KEY: exon
? LOCATION: 30470..30626
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 13
? FEATURE:
? NAME/KEY: exon
? LOCATION: 30187..30274
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 14
? FEATURE:
? NAME/KEY: exon
? LOCATION: 29945..30073
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 15
? FEATURE:
? NAME/KEY: exon
? LOCATION: 34092..34358
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 2
? FEATURE:
? NAME/KEY: exon
? LOCATION: 33779..33963
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 3
? FEATURE:
? NAME/KEY: exon
? LOCATION: 33493..33591
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 4
? FEATURE:
? NAME/KEY: exon
? LOCATION: 33297..33408
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 5
? FEATURE:
? NAME/KEY: exon
? LOCATION: 32959..33094
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 6
? FEATURE:
? NAME/KEY: exon
? LOCATION: 32569..32628
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 7
? FEATURE:
? NAME/KEY: exon
? LOCATION: 32386..32468
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 8
? FEATURE:
? NAME/KEY: exon
? LOCATION: 31894..32080
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 9
? FEATURE:
? NAME/KEY: exon
? LOCATION: 31363..31534
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 10
? FEATURE:
? NAME/KEY: exon
? LOCATION: 31131..31284
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 11
? FEATURE:
? NAME/KEY: exon
? LOCATION: 30816..31011
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 12
? FEATURE:
? NAME/KEY: exon
? LOCATION: 30470..30626
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 13
? FEATURE:
? NAME/KEY: exon
? LOCATION: 30187..30274
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 14
? FEATURE:
? NAME/KEY: exon
? LOCATION: 29945..30073
? OTHER INFORMATION: /gene= "AR"
? OTHER INFORMATION: /number= 15
```



```

: OTHER INFORMATION: /gene= "AR"
: OTHER INFORMATION: /number= 15
:
: FEATURE:
:
: NAME/KEY: exon
: LOCATION: complement (29664..29856)
: OTHER INFORMATION: /gene= "ARS"
: OTHER INFORMATION: /number= 16
:
US-08-814-095-7

```

Alignment Scores:	
Pred. No.:	0.0215
Score:	165.00
Percent Similarity:	29.76%
Best Local Similarity:	21.78%
Query Match:	3.36%
DB:	3
Gaps:	33
Length:	35060
Matches:	143
Conservative:	57
Mismatches:	168
Indels:	305

44-720-086-5 (1-908) X US-08-814-095-7 (1-35060)

OY	45	ValGlyArgProGluTyrGlySerArgLysHisProProValGlnSerSerAspThrProlys	64
Db	24360	TTCGGCTCCGGTGCGAGMAAGCAGCGGGGTCCATGGTGCGCTCCGA-----	24313
OY	65	AspProAlaValThrThrylserGlnPrometAlaGlnAspserGly-ProSerAspLeu	84
Db	24312	-----AAGGGATGCCAGAAACAGAGACAGCGCCCCCGGGGGTC	24271
OY	85	-----LeuProAsnGlyAspIleuGluLysArgSerGluProGlnProGluGluGly	101
Db	24270	TTCAGCGCAATGCC-----CGCAGCGCGCCCCCACGCACCTGCACC	24229
OY	102	SerProAlaAlaGlyGlnLysGlyGlyAlaProAlaGluGluGlyGlyThrGluThrPro	121
Db	24228	AGCAGCTCGCATCTCCCGCCCTCAGCCCC-----ACTCCT	24190
OY	122	ProGluAlaSerArg-----AlaValGluAsnGlyCybCyS---	133
Db	24189	CCACCAGAGAGCCAGAGAGAGAGAAGAGATGGGAAAGCCAGGAAAGCGCTGTGCAGC	24130
OY	134	-----ValThrLysGluGlyArgGlyAlaSerAlaGlyGlu-----	145
Db	24129	AGACACTGCGGGGCCCTCATGGCTCCAGGCGAGCGCGCTGTGCGGAGAAAAGAGG	24070
OY	146	-----GlyLysGluGlnLysGlnThrAsnIleGluSerMetLysMetGlu	160
Db	24069	AAAGGTGAAGGTCTGGCTCGAGAACGACAGAGAGAGAGATTATGGGCACCTGTGCGCCCA	24010
OY	161	Gly-----SerArgLysArgLeuArgGlyGlyLeuGlyTyrGluSerSerLeu	176
Db	24009	GGGTATCGCTCAGCTGCAGAGGA-----GGTGGGCGAGGTGGCGAAGA-----	23965
OY	177	ArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTyrTrpIleSerLys	196
Db	23964	-----TGAGGGGAG	23956
OY	197	ArgLysArgAspGluTyrPleuAlaArgTyrPheArgGluAlaGluLysAlaLysVal	216
Db	23955	AGAGGGAGGGAG-----GGAAGGAGACAGCGCAGAGACAAGG-----	23920
OY	217	IleAlaValMetAsnAlaValGluGluAsnGlnAlaSerGlyGluSerGlnLysValGlu	236
Db	23919	-----ACACAGGCCACCGCGCAGAGAGGGGAAAGAGATC	23887
OY	237	GluAlaSerProProAlaValGln---GlnProThrAsp-----Pro	249
Db	23886	AGGCACACACCCATATCCAGACAAATGACAGGAACAGAGAGCAACCTTTCTGCTCCG	23822
OY	250	AlaSerProThrValAlaThrThr--ProGluProValGlyGlyAspAlaGlyAspLysAs	269
Db	23826	CAGGGGTCCACTGCCCTCACCCTTCTCCGGAACGTGGCGCGGCCACCTCGTCGACGTAG	23767
OY	269	nlaIthrLysAlaAlaAspAspGluProGluTyrGluLysArgLysArgGlyPheGlyIleGly	289

Db	23766	GGCACCAGAAAGCTCAGTCCACGAACAAGCCACCGGACACGAGGCGCGGCTCACCGACGG	23707
Qy	289	YGIuLeuValTrpGIuLyLeuArg-----GIyheSerTrpTrpProGIuYArgII	306
Db	23706	GCAGCGCGCGCTGGGGAGCGTTGGGGCGCTCGAAGGGCGGATCC-----CC	23662
Qy	306	eValSerTrpTrpMetThrGIuArgSerArgAlaIaGIuGIYThrArgTrpValMetTr	326
Db	23661	GCTCTCTCTCG-----GTGGCCACAGACCCGG--GGCCTGGGGGGCGGTGGGCTGGGCA	23611
Qy	326	pPheGIYAspGIuYsPhe-----SerValValCysValGI	338
Db	23610	TTCAAGGGGTGGAAAGCTCAGGTCGCCGGGGGTCCCCCAATTCCTGACATGTCACATCCCA	23551
Qy	338	uYsLeuMetProLeuSerSerPheCysSerAlaPheHisGIuAlaThrTYrAsnIysGI	358
Db	23550	CCCCAACCTGGCCCTTCTCATCCATGT-----	23523
Qy	358	nPrometTYrArgYsAlaIeTYrGIuValLeuGIuValAlaSerSerArgAlaGIY	378
Db	23522	-----CGGCACCGAGCGCTGATGCCG	23500
Qy	378	sLeuPhePro--AlaCysHisAspSerAspGIuSerAspSerGIYsAlaIaGIuVa	397
Db	23499	CGGGTTCCTCATGACGTCCGACGC--TCCACAGCGAGCCCCACGAACAAGCT-----	23449
Qy	397	IGIuAsnIysGIuMetIleGIuTrpAlaLeuGIYGIYpHeGIuProSerGIYProIysGI	417
Db	23448	-----CGGCAGCGCGCTCCCCACGAGAC	23426
Qy	417	YLeuGIuProProGIuGIuGIuIYsAsnProTYrIYsGIuValTYrThraSPMetTrpVa	437
Db	23425	TCGAGAGCCCCCATCGCTGAGAGCTCCG-----	23395
Qy	437	IGIuProGIuAlaAlaAlaTYrAlaProProProAlaIaLYsProArgLYsSerTh	457
Db	23394	-----TCACAGCGCGCGCGCGGAG-----	23374
Qy	457	rThrGIuYsProLYsValIYsGIuIleIleAspGIuArgThrArgGIuArgLeuValTY	477
Db	23373	-----GSCAGTAGAGACCGAGAAAGSTC-----	23350
Qy	477	rGIuValArgGIuLYsCysArgAsnIleGIuAspIleCysIleSerCysGIYSerLeuAs	497
Db	23350	-----	23350
Qy	497	nValThrLeuGIuHisProLeuPheIleGIYGIuMetCysGIuAsnCYsLYsAsnCYsPh	517
Db	23350	-----	23350
Qy	517	eLeuGIuCYsAlaTYrGIuTYrAspAspArgIYTYrGIuSerTYrCYsThrIleCYsCY	537
Db	23350	-----	23350
Qy	537	sGIYGIuArgGIuValLeuMetCysGIYAsnAsnAsnCYsArgCYsPheCysValGI	557
Db	23349	---GGCGCTCCCGCGCCCGAGCCGAGAGCCGGGAATTGGGAGG-----	23308
Qy	557	uCYsValAspLeuLeuValGIYProGIY-----AlaIaGIuAlaIaIeIYsGI	574
Db	23307	-----GGGCTGGGGGCTCGAGGGCTTTCGACAGCGCTGAAGGG	23267
Qy	574	uAspProTrpAsnCYsTYrMetCysGIYHisLYsGIYThrTYrGIYLeuLeuArgArg	594
Db	23266	CGAGCG-----GGACTCTCGCTGCCG	23243
Qy	594	gGIuAspTrpProSerArgLeuGIu-----MetPhePheAlaAsnAsn-----	608
Db	23242	GGACTCCGAGATCCCGCGCTCGACAGGGCGGGCGCGCTATTTTGGGGCGACAGAGCTGCTA	23183
Qy	609	-----HisArgGIuLeuPheAspProProLYsVal--TYrProProValPr	623
Db	23182	CGGTTCCGGGATCTCCAGCCAAAGGATCAAAATTCCCGGGGTGGGAACCTCCCGCTCTCC	23123





Db 2675 GACGGGCAAGATGATCCCTAGGTGTCTCCGGTGGGG-----GATCCCAAGAT 2625  
Qy 302 TrpProGlyArgIleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGluGlyThr 321  
Db 2624 CTTCCCTCCAG-----AGCACACAGGAGTACTACTCGTGGG--- 2583  
Qy 322 ArgTrpValMetTrpPheGlyAspGlyLysPhe-----SerValValCysVal 337  
Db 2582 -----GGCATGTCTCTAGCTACCGGCCCTCCCAATGTG 2547  
Qy 338 GluLysLeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLys 357  
Db 2546 GCCAACTCTCTCCCTAGCTCTGGCTGGGCCAACACA----- 2511  
Qy 358 GlnProMetTyrArgLysAlaIleTyrGluValLeuGlnValAlaLysSerArgAlaGly 377  
Db 2510 -----CAGTAAGTAAGGTGTATCGGAATGGGTGGGT 2478  
Qy 378 LysLeuPheProAlaCysHisAspSerAspGluSerAspSerGlyLysAlaValGluVal 397  
Db 2477 ATGGGAGCATGCTGGCTCAGCTCGGGGGCGGTGGGACAGCGCCAGGAGGCCAGC 2418  
Qy 398 GlnAsnLysGlnMetIleGluTrpAlaLeuGlyGlyPheGlnPro---SerGlyProLys 416  
Db 2417 CTCAAGGGCATCTCTGTCATGTGGAAAGATGAATTAGGCTGGGACACACCCGCC 2358  
Qy 417 GlyLeuGluProProGluGluGluLysAsnProTyrLysGluValTyrThrAspMetTrp 436  
Db 2357 CACCTACGGGTGTATCCCAATGCCCGCGCGCACACTCAG----- 2316  
Qy 437 ValGluProGluAlaAlaTyrAlaProProProProAlaLys----- 451  
Db 2315 -----GCCCGACCATCTGAGCCCTCTCTGAGAGGCGGACGACAGACA 2271  
Qy 452 LysProArgLysSerThrThrGluLysProLysValLysGluIleAlaAspGluArgThr 471  
Db 2270 GGGCGAGGGGAGTGGAGATGGAAACAGGAGGGGCTCAGACCATG-----AAACA 2217  
Qy 472 ArgGluArg-LeuVal-----TyrGluValArgGlnLysCysAspGAs 485  
Db 2216 TTGAGAAATCTTAGCGGACTTGGCTGATCCATACAGATATCGT---AGTGTCTAGA 2160  
Qy 485 nileGluaspileCysIle---SerCysGlySerLeuAsnValThrLeuGluHisProLe 504  
Db 2159 TAGCTTATGTTTTCATCTAATTTTGTGGGGGTGTGCTGTGTGCTGTGTGTGTGT 2100  
Qy 504 uPheIleGlyGly---MetCysGlnAsnCysLysAsnCysPheLeuGluCysAlaTyrG1 523  
Db 2099 GCAGGTGGGTGCTTGGCTGTGCAACGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040  
Qy 523 nTyrAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyGlyArgGluValLe 543  
Db 2039 TCATGTATGTGCTGCGCATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1980  
Qy 543 uMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysValAspLeuVal 563  
Db 1979 TGTGTGCTGTGTTTCATCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1920  
Qy 563 lGlyProGlyAlaAlaGlnAlaAlaLysGluAspProTrpAsnCysTyrMetCysG1 583  
Db 1919 TGCT-----GTGTTTCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1893  
Qy 583 YHis 584  
Db 1892 GCAT 1889  
RESULT 6  
US-08-458-568A-11  
; Sequence 11, Application US/08458568A  
; Patent No. 5821339  
; GENERAL INFORMATION:  
; APPLICANT: Schaffer, Priscilla A.  
; APPLICANT: Yeh, Lily

TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus  
TITLE OF INVENTION: Infections  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339ris  
STREET: One Liberty Place, 46th floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,568A  
FILING DATE: 02-JUNE-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/065,146  
FILING DATE: 05-MAY-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Leary Ph.D., Kathryn R.  
REGISTRATION NUMBER: 36,317  
REFERENCE/DOCKET NUMBER: DFCI-0029  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12001 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Herpes simplex virus  
STRAIN: Herpes Simplex Virus Type 1  
US-08-458-568A-11  
Alignment Scores:  
Pred. No.: 0.00942 Length: 12001  
Score: 161.00 Matches: 117  
Percent Similarity: 28.93% Conservative: 45  
Best Local Similarity: 20.89% Mismatches: 199  
Query Match: 3.28% Indels: 199  
DB: 1 Gaps: 28  
US-09-720-086-5 (1-908) x US-08-458-568A-11 (1-12001)  
Qy 2 ProSerSerGlyProGlyAspThrSerSerSerLeuGluArgGluAspAspArgLys 21  
Db 5901 CCCGGCGCGGGCGCGGT-----CGCCGCGCGCGGTCTCGCTG 5936  
Qy 22 GluGlyGluGluGlnGluGluAsnArgGlyLysGluGluArgGlnGluProSerAlaThr 41  
Db 5937 CGTCGTCTACTGTCCTCCCTCATCTCGTCTCGTCTCGAAAGCGG---GGGTCCGGGGCG 5993  
Qy 42 AlaArgLysValGlyArgProGlyArgLysArgLysHisProProValGluSerSerAsp 61  
Db 5994 GCGAGCGCGGGCGTCTCGGCGTCTCGGATCTCGGACGCGCTCTCTACCATGG----- 6047  
Qy 62 ThrProLysAspProAlaValThrThrLysSerGlnPro-----Met 75  
Db 6048 AGGCCAGCAGACCCAGCTGTCCGGCGGACGCGCGTCTCGCGCGCGCTCG 6107  
Qy 76 AlaGlnAspSerGlyProSerAspLeuLeuProAsnGlyAspLeuGluLysArgSer--- 94  
Db 6108 TGCCCGCGCGGGGGCCCTCC-----CGTCCCGCGCGGGCGTCTCGAGGTCTGTGGGGT 6161

```

QY 95 -----GluProGlnProGluGluGly 101
Db 6162 GGTGGGGGTGTGTGGGGGTCTCCCGCTCTCCCTCTCCGCGCCCAACCGGAGGG 6221
QY 102 SerPro-----AlaAlaGlyGlnGlyGlyAlaPro 112
Db 6222 CCCCCCTCTCTCCGCTGTGGCTCGGGGTGGGGCGCGCGCTCGTGGGCGCGGG 6281
QY 113 AlaGluGlyGluGlyThrGluThrProProGluAlaSerArgAlaValaGluAnGlyCys 132
Db 6282 AGCGGGGGCTCTCTCTCTCTCCGACCGCATCGCGCATGGGGCGCATCTCCGGGATA 6341
QY 133 CysVal-----ThrGlyGluGlyArgGlyAlaSerAlaGlyGlu--GlyGlyGlu 148
Db 6342 CGGCTGGACGGCGGACGTAAGACAGTACCTACCGACTCTCGATGGGGGAGAGGG 6401
QY 149 GlnGlySerGlnThrAsnIleGluSerMetGluGlySerArgGlyArgLeuArgGly 168
Db 6402 GCGAGACCCACGACCGCCGACGACCGCGCTCGACGCGGACGCGGACGCGGACCGGTCG 6461
QY 169 GlyLeuGlyTyrGluSerSerLeu--ArgGlnArgProMetProArgLeuThrPheGln 187
Db 6462 ATGCTTGGGTGGGAAAAAGACAGGAGCGCGCATCCCTCCGCGCTTCTCTC----- 6515
QY 188 AlaGlyAspProTyrTyrIleSerLysArgLysArgAspGlu----- 201
Db 6516 -----CGCTATCGCGCTCCCGCGCGGCGGCGCTGACGCTGCTGCTC 6560
QY 202 TyrLeuAlaArg-----Tyr-----LysArgGlu 209
Db 6561 TGGGGGTCTCCGCTCGGCTGTGATCCGTGCGGACCGCGCTCCGTGTGACGATCG 6620
QY 210 AlaGluLysLeuAlaValIleAlaValMetAsnAlaValaGluGluAnGlnAlaSer 229
Db 6621 GGGGCTCTCCGCTCTATATAGTC-----CCA 6647
QY 230 GlyGluSerGlnLysValaGluGluAlaSerProProAlaValaGlnGlnProThrAspPro 249
Db 6648 GGGGCGGCGGAGAGAGAGACGAGCGGCGCGCGC-----CCCCCGCCCCC 6698
QY 250 AlaSerProThr-----ValAlaThrThr----- 257
Db 6699 GCGGCGCCACCGGACGGAATTCATATGACGACCGCGCGCGGACGCGGACGCGC 6758
QY 258 -----ProGluProValaGlyGlyAspAlaGlyAspLysAlaThrLysAla 273
Db 6759 GGGGCGCGGCGCGCGCGCTGTGTCGAAACCGCGCGCGCGCAT----- 6806
QY 274 AlaAspAspGluProGluTyrGluAspGlyArgGlyPheGlyIleGlyGluLeuValTyr 293
Db 6807 -----CCGCGCATGTGCA----- 6821
QY 294 GlyLysLeuArgLysPheSerTyrProGlyArgIleValSerTyrTyrMetThrGly 313
Db 6822 -----TGCGCGGCGCGC-----GAGCGCGGTGGGTCCGCGCGC 6854
QY 314 ArgSerArgAlaAlaGluGlyThrArgTyrValMetTyrPheGlyAspGlyLysPheSer 333
Db 6855 CCGCGCGCATGCGCATTCATACCGC----- 6881
QY 334 ValValCysValaGluLysLeuMetProLeuSerSerPheCysSerAlaPheHisGlnAla 353
Db 6882 -----CCGATCGGCGGCTTCGCGCTTCGCGCTTCGCGCATGCTTA 6917
QY 354 ThrTyrAsnLysGlnProMetTyrArgLysAlaIleTyrGluValaGluValaLaser 373
Db 6918 ACGAGAAACGGGCAAG-----GGG 6935
QY 374 SerArgAlaGlyLysLeuPheProAlaCysHisAspSerAspGluSerAspSerGlyLys 393
Db 6936 GCGGCGCGCGCGCGCGCATTTCC-----GGTTTCG 6965

```

```

QY 394 AlaValaGluValaGluAnLysGlnMetIleGluTyrAlaLeuGlyGlyPheGlnProSer 413
Db 6966 GCGGTAAATG-----AGATRCGAGCCCGC 6989
QY 414 GlyProLysGlyLeuGluProProGluGluGluLysAsnProTyrLysGluValaTyrThr 433
Db 6990 GCGCGCGTGGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7031
QY 434 AspMetTyrValaGluProGluAla--AlaAlaTyrAlaProPro-----ProProAlaLys 451
Db 7032 GACGCGGAGCAACGAGGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 7091
QY 451 LysProArgLysSerThrThrGluLysProLysValaLysGluIleLeuAspGluArgTh 471
Db 7092 TTGGCGCGCGCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7151
QY 471 Arg-----GluArgLeuValaTyrGluValaArgLysGlyCysArgAsn 485
Db 7152 GCGAAGCTTCGCACTTCTCCCATATATATATATATATATATATATAGGCGGAGTGCAGCAC 7209

RESULT 7
US-09-154-750A-75
: Sequence 75, Application US/09154750A
: Patent No. 6432640
: GENERAL INFORMATION:
: APPLICANT: Vogelstein, Bert
: APPLICANT: Kinzler, Kenneth
: APPLICANT: Polyak, Kornelia
: TITLE OF INVENTION: p53-Induced Apoptosis
: FILE REFERENCE: 1107.75357
: CURRENT APPLICATION NUMBER: US/09/154, 750A
: CURRENT FILING DATE: 1998-09-17
: PRIOR APPLICATION NUMBER: 60/059,153
: PRIOR FILING DATE: 1997-09-17
: PRIOR APPLICATION NUMBER: 60/079817
: PRIOR FILING DATE: 1998-03-30
: NUMBER OF SEQ ID NOS: 93
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 75
: LENGTH: 2608
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-154-750A-75

Alignment Scores:
Pred. No.: 0.00132 Length: 2608
Score: 159.50 Matches: 153
Percent Similarity: 32.04% Conservative: 79
Best Local Similarity: 21.13% Mismatches: 278
Query Match: 3.25% Indels: 217
Gaps: 33
DB:

US-09-720-086-5 (1-908) x US-09-154-750A-75 (1-2608)
QY 4 SerGlyProGlyAspThrSerSerSerSerLeuGluArgLysAspArgLysGluGly 23
Db 632 TCTGCCCAAGC-----TGCGAGACCATCAACGACCATGAAAGG 670
QY 24 GluGlu-----GlnGluGluAsnArgLysGluGluArgLysGluGluPro 38
Db 671 GAGGATCTCGGAACCTGCACTGAGGCTGATGACACGAGAGATGGCGGTGAAGCCCTTGA 730
QY 39 SerAlaThrAlaArgLysValaGlyArgProGlyArgLysArgLysHisProProValaGlu 58
Db 731 GTCGAGAGAGAGAGAGAGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 790
QY 59 SerSerAspThrProLysAspProAla--ValThrThrLysSerGlnProMetAlaGln 77
Db 791 CAATCAAGAAATCCAGAGACTCCAGCGCCAGCAAGAGCAAGAGAGAGAGAGAGAGAGAG 850
QY 78 Asp-----SerGlyProSerAspLeuLeuProAsnGlyAspLeuGluLysArgSerGly 95
Db 851 GATTAAAGATCTGAGAGCAAGAGCTGTCTCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 910

```

QY 96 ProGlnProGluGluGly---SerProAlaAlaGlyGlnLysGlyGlyAlaProAlaGlu 114  
 Db 911 CATGAAGTCTGAGCTGTGTACCGCTCCCTAGGCTGGAAACGGGAGCTGGAGCAGCTGGGGA 970  
 QY 115 GlyGluGlyThrGluThrProGluAlaSerArgAlaValIleGluAsnGlyCysCysVal 134  
 Db 971 GGAGAGCGCACTCGGGAGATGAGAGAGACCAACGGCTGTCTCCAGGA----- 1018  
 QY 135 ThrLysGluGlyArgGlyAlaSerAlaGlyGlyLysGluGlnLysGlnThrAsnIle 154  
 Db 1019 ---AGAGCTGGAAGGCTGAGAGAGCTGGGGCGCCAGAGAGAAAGTGCAGAGAGCGCT 1075  
 QY 154 eGluSerMetLysMetGluGlySerArg-----GlyArgLeu-ArgGlyGlyLeuGlyT 172  
 Db 1076 GGTGGCTTGGAGCTGGAGAACGAGAGGCTGTGGCCAAAGTGCAGAGACTGGGAGACT 1135  
 QY 172 rpGlu-----SerSerLeuArgGlnArgProMetProArgLeuThrPheG 187  
 Db 1136 GGACAGACCATGGGCTGAGCATCAGGACTCCAGAGACTTTCCA----- 1182  
 QY 187 InAlaGlyAspProTyrrIleSerLysArgLysArgAspGluTrpLeuAlaArgTrpL 207  
 Db 1183 -----GATTCGTGTTGAGCTG-CAGC 1203  
 QY 207 ysArgGluAlaGluLysLysAlaLysValIleAlaValMetAsnAlaValGluGluAsnG 227  
 Db 1204 AGAGGGAGCTTGCCTTGAAGGACAAAGACAGCGCCGTCAC-----CAGCAGCGCC 1253  
 QY 227 InAlaSerGlyGluSerGlnLysValGluLysValGluLysValGluLysValGluLys 247  
 Db 1254 CGGGGGTGGAGAGAGCGCAGCAGCTGCGAGGAGGCTCCGGAGGTCAGCGGCGCAG 1313  
 QY 247 hrAspProAlaSerProThrValAlaThrProGluProValGlyGlyAspAlaGlyA 267  
 Db 1314 CT-----GTTGGAGGAGAGAGAGAG 1334  
 QY 267 spLysAsnAlaThrLysAlaAlaAspAspGluProGluThrGluAspGlyArgGlyPheG 287  
 Db 1335 CGCAGACCCAGCAGCGCTGCGCGGAGGCTCCAGAAACGGGTCTGCTGTCTACCAAG 1394  
 QY 287 lyIleGlyGluLeuValTrpGlyLysLeuArgGlyPheSerTrpTrpPro----- 303  
 Db 1395 GAGCG-----GGACGGTATGCGGGCCA-TCCTGGGGTCTCTACGACAGCCA 1438  
 QY 304 -----GlyArgIleValSerT 309  
 Db 1439 GCTGACCCCGCGGAGTACTACCCAGCTGACCGGCGCATGCGGGAGCTTGAGGATAT 1498  
 QY 309 rp-----TrpM 311  
 Db 1499 GGTGCAGAGGTGCACAGCCAGCCGAGATGAGGCTCAGCTGTGCGAGCGCCCTGA 1558  
 QY 311 etThrGlyArgSerArgAlaAlaGluGlyThrArgTrpValMetTrpPheGlyAspGlyL 331  
 Db 1559 GGAGCTGGAGGCGGAGAAACAAAGAGCAGACATGCTCG-----A 1597  
 QY 331 ysPheSerValValCys-----ValGluLysLeuMetProLeuSerSerPheCysSerA 349  
 Db 1598 GATGGAGCTGAAGATGCTGAAGTCTCAGTCCAGCTCTGCGGAAACAGAGCTTCTCTGTC 1657  
 QY 349 laPheHisGlnAlaThrTyrrAsn----- 356  
 Db 1658 CAGGAGGAGGCGGACACGCTCAGGTTGAAGGTTCGAGGAGCTGGAAGCGGAGGAGTCG 1717  
 QY 357 -----LysGlnProMetTyrrArgLysAla---IleTyrrGluValLeuGlnValA 372  
 Db 1718 GCTGGAGGAGGAGGAGGATGCTGGAGGACAGCTGGAGCGGAGCTCTGCAGGGTGA 1777  
 QY 372 laSerSerArgAlaGlyLysLeuPheProAlaCys-HisAspSerAspSer 391  
 Db 1778 CTATGACAGCAGAGGAGCA-----AAGTGTGTGACATGAGCTGAAACCCACAGTGT 1831

QY 392 GlyLysAlaValGluValGlnAsnLysGlnMetIleGluTrpAlaLeuGlyGlyPheGln 411  
 Db 1832 GGCAGGAGCGCTGCGCGAGGAGCAGACCA-----GCTCAGCGGAGGAGTCGA 1882  
 QY 412 ProSerGlyProLysGlyLeuGluProProGluGluGluLysAsnProTyrrLysGluVal 431  
 Db 1883 GCGACTGCGCGGCTCTCTGCGCGCATCGAGAGAGGAGCAGCCTCCAGCCGA----- 1936  
 QY 432 TyrrThrAspMetTrpValGluPro-GluAlaAlaAlaTyrrAlaProProProPro 449  
 Db 1937 -----CCTTGAAGGCTGCCCGCGGAGTCTGCCATCGCTCCCAAGGA 1975  
 QY 450 -----AlaLysLysProArgLysSerThrThrGluLysProLysValLysG 465  
 Db 1976 GTGGGAGAGCTGAAGAGCAGGTGGAGTGCAGAGTGCAGAGTGAAGAACCCAGGCTCAAGGA 2035  
 QY 465 uIleIleAspGluArgThrArgGluArgLeuValTyrrGluValArgGlnLysCysArgAs 485  
 Db 2036 GGTTCACAGACCAAG-----ATCCAGGAGTTCGCAAGGCTGTCTACAC 2080  
 QY 485 nIleGluAspIleCysIleSerCysGlySerLeuAsn-----Valth 499  
 Db 2081 GCTACCGGCTACAGATCGACATCAGCAGGAGAACAGTACCGGCTGACCTGCTGTA 2140  
 QY 499 rLeuGluHisProLeuPheIleGlyGlyMetCysGlnAsnCysLys-----AsnCy 516  
 Db 2141 CGCCGAGCAGCA-----GCGGACTGCTATCTTCAGGCCACCGCCCTCG 2188  
 QY 516 sPheLeuGluCysAlaTyrr-----GlnTyrrAspAspGlyTyrrGlnSerTy-CysTh 534  
 Db 2189 GGTTCAGATGAGCTACTGAGAGACAGAGTCTCAGCAGCGTGGGCGAGCTCATCGAG 2248  
 QY 534 rIleCysCysGlyArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPh 554  
 Db 2249 GTGACCTCGCGCGCAGGAGCAGTCCCTGCTCTCTCAGCTCGC----- 2294  
 QY 554 eCysValGluCysValAspLeuValGlyProGlyAlaAlaGlnAlaAlaIleLysG 574  
 Db 2295 -----TCACCTCGAGGCTCTTTCAGCGCC----- 2318  
 QY 574 uAspProTrpAsnCysTyrrMetCys---GlyHisLysGlyThrTyrrGlyLeuArgAr 593  
 Db 2319 -AGACCGGCGGCTGAGCTGAGCTCGGCGGCGCATAGC-----CGAG 2359  
 QY 593 gArgGluAspTrpProSerArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPh 613  
 Db 2360 CCACTCTGTTGGCTGAC---CTGACAGTCT----- 2387  
 QY 613 eAspProLysValTyrrProProValProAlaGluLysArgLysProIleArgValLe 633  
 Db 2388 -----CCCTGCGCGCGCAGCAGCAGGCTGGGTGCAGCTCTCTGC 2425  
 QY 633 uSer 634  
 Db 2426 CTCT 2429

RESULT 8  
 US-08-061-376-4  
 ; Sequence 4, Application US/08061376  
 ; Patent No. 6175000  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Evans, Glen A.  
 ; APPLICANT: Djabali, Malek  
 ; APPLICANT: Salleri, Licia  
 ; APPLICANT: Parry, Pauline  
 ; TITLE OF INVENTION: CHARACTERIZATION OF A CHROMOSOME 11Q23  
 ; NUMBER OF SEQUENCES: 12  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark  
 ; STREET: 444 South Flower Street, Suite 2000  
 ; CITY: Los Angeles  
 ; STATE: California





```
Db 4008 AGGTCCAGAGCAGCAGCAAAACAGAAAAAGTGGCTCCCGCCCAAGTATCCCTGTAAACA 4067
Qy 459 uLySerProLyValLyGlu-
Db 4068 AAACCAAAAGAAAGAAAGAAACCAACCTCCGGTCAATAAGCAGGAGATGCAGGCACITTT 4127
Qy 466 ----llelleaspGluArgThrArg-
Db 4128 GAACATCCCTCAGACACTCTCCCAATGCCAATAGTCTTAAGCAAAAAATTCAGCAGATGG 4187
Qy 473 ----GluArgLeuValTyrgluValArgGlnLyCys-----ArgAsnilleGluasp-- 488
Db 4188 AGTCCACAGGATCAGATGCAGTCTTAAGGAGGATTTGAAGCAGAAAAATGTGTGGAGAT 4247
Qy 489 -----
Db 4248 GGGAGGCTTAGGAATCTTGACTTCTGTCTTATACACCCAGGGTGTGTTGCTTCTCTG 4307
Qy 493 sglySerLeuAsnValThrLeuGluHisProLeuPheilleGlyGlyMetCysGlnAsnC 513
Db 4308 TGCCAGTAGTGGG-----CATGTAGAGTTTGTG-----TATTGCCAAGTCTG 4349
Qy 513 s-----LysAsnCysPheLeuGluCysAlaTyrglnTyrglnTyrglnTyrglnTyrgln 528
Db 4350 TTGTGAGCCCTCCCAAGTTTGTGTTAGGAGAACGAGCGCCCTCTGGAGGACAGCT 4409
Qy 528 yTyrglnSerTyrglnTyrglnTyrglnTyrglnTyrglnTyrglnTyrglnTyrglnTyrgln 538
Db 4410 GGAATAATGGTGTGTGCTGCTGCTGCAAAATCTGTACGTTTGTGGAGGCAACATCAGGC 4469
Qy 539 -GlyArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluC 558
Db 4470 TACAAGAGCAGCTGCTGGAGTGT-----AATAAGTGGCAACACAGCTATCACCTTGAGTG 4523
Qy 558 sValaspLeuLeuValGlyProGly-----AlaAlaGlnAlaAlaAlaAlaAlaAlaAla 577
Db 4524 C-----CTGGACCAAACTACCCCAACCAACCAACCAACCAACCAACCAACCAACCAAC 4571
Qy 577 pAsnCysTyrglnMetCysGlyHisLyGlyThrTyrglnTyrglnTyrglnTyrglnTyrgln 597
Db 4572 GATCTGTACCAAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4631
Qy 597 pProserArgLeuGlnMetPhePheAlaAsnAsnHisasp 610
Db 4632 GATGCACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4671

RESULT 9
US-08-320-559-1
; Sequence 1, Application US/08320559
; Parent No. 5633135
; GENERAL INFORMATION:
; APPLICANT: Croce, Carlo
; TITLE OF INVENTION: Diagnostics, Therapeutics and Methods for
; TITLE OF INVENTION: Detection and Treatment of Acute Leukemias
; TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the
; TITLE OF INVENTION: All-1 Region
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESS: Woodcock Washburn Kurtz Mackiewicz & No. 5633135ris
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/320,559
```

```

; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/062,443
; FILING DATE: 14 MAY 1993
; PRIOR APPLICATION DATA: US/07/971,094
; APPLICATION NUMBER:
; FILING DATE: 30-OCT-92
; PRIOR APPLICATION DATA: US/07/888,830
; APPLICATION NUMBER:
; FILING DATE: 27-MAY-92
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/805,093
; FILING DATE: 11-DEC-91
; ATTORNEY/AGENT INFORMATION:
; NAME: Deluca, Mark
; REGISTRATION NUMBER: 33,229
; REFERENCE/DOCKET NUMBER: TJU-0855
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14255
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; ANTI-SENSE: No
; US-08-320-559-1

Alignment Scores:
Pred. No.: 0.0195 Length: 14255
Score: 158.50 Matches: 148
Percent Similarity: 32.17% Conservative: 101
Best Local Similarity: 19.12% Mismatches: 275
Query Match: 3.23% Indels: 254
DB: 1 Gaps: 33

US-09-720-086-5 (1-308) x US-08-320-559-1 (1-14255)
Qy 10 SerSerSerSerLeuGluArgGluAspArgLyValGluGluGluGluGluGluGluGlu 29
Db 2404 GCTGACAAAGAGCGTGGAGAGGACAGAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2463
Qy 30 ArgGlyLyGluGluArgGlnGlu-
Db 2464 AATAAGCGGGAGTCAAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2523
Qy 41 ThrAlaArgLyValGlyArgProGlyArgLyArgLyArgLyArgLyArgLyArgLyArg 59
Db 2524 GCTTTGTATCTCTGGGTAGGGTTTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2583
Qy 60 SerAspThrProLyAspProAlaValThrThrLySerGlnProMetAlaGlnAspSer 79
Db 2584 TCATCTCTGCCAAA-----AAAGCAACAGCGCGGAGAGAGAGAGAGAGAGAGAGAG 2637
Qy 80 GlyPro-----SerAspLeuLeuProAsnGlyAspLeuGluLys 92
Db 2638 GGGACTGATATTACTTCTGTGACTCTTGGGATACACAGCTGT-CAAAACCAAAATACT 2696
Qy 93 ArgSerGluProGlnProGluGluGlySerProAlaAlaGlyGlnLyGlyGlyAlaPro 112
Db 2697 TATAAGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2756
Qy 113 AlaGluGlyGluGlyThrGluThrProPro-----
Db 2757 ATCCCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 122
Qy 123 -----GluAlaSerArgAlaValGlu 129
Db 2817 TTCCACTTCTCCATAGGCTCCATGTTGGCTCAGCAGACAGCTTCCAACTACTGACAA 2876
Qy 130 AsnGlyCysCysValThrLyGluGlyArgGlyAla----- 141
```



Db 2877 GAGGGTTGCCAGCTCTCTAAAGGCAAGCTCAGCTCTGCAAGATTGAGAAAGCTAA 2936  
OY 142 -----Sera1aglygluGlyLeu-GluGlnLysGlnTh 152  
Db 2937 GAGTCTTAAACAAACCAAGCCCAAGGCAAGGCTCAAGAAAGTACTCATAGAGAC 2996  
OY 152 rAsn1leG1userMetLysMetGluGlySerArgGlyVarGLeuArgGlyGlyLeuGlyTr 172  
Db 2997 CTCTGTGCAGAGACCCCGGATTAAACATGTCTGCAGAAAGAGAGAGCTGTTGCCCG 3056  
OY 172 pG1userSerLeuArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTy 192  
Db 3057 AAAACAGAGCTGTCTTCTGATGACATGCCCACTGAGT----- 3096  
OY 192 rTyrl1eSerLysArgLysArgAspGluTrpLeuAlaArgTrpLysArgGlyAlaGly 212  
3097 -----GCCTTACCATTG-----GAAAGA 3113  
OY 212 sLysAlaLysVal1leAlaValMetAsnAlaValGluGlnAlaSerGlyGlyLys 232  
Db 3114 ACGAGAAAGATTGTTCTTCTTCAATGGGAGATGATGACAGATCATCAATTGCTGCTCA-- 3171  
OY 232 rGlnLysVal1GluGlnLysSerProProAlaVal1GlnGlnProThrAspProAlaSerPr 252  
Db 3172 -----GAAAGTGTGAACCTTGTCT-----CCACCATCAACCAATTAAAC 3215  
OY 252 OThrValAlaThrThrProGluProVal1GlyLysAlaGlyAspLysAsnAlaThrLys 272  
Db 3216 TGTC-----ACTAGAAA 3227  
OY 272 sAlaAlaAspAspGluProGluTrpGluAspGlyArgGlyPheGly1leGlyGlnLeuVal 292  
Db 3228 CAAGGACCCCAAGAACCTCCAGTAAGAAAGAGAGCTG-ATC-----GAGCGGT 3277  
OY 292 1TrpGlyLysLeuArgGlyPheSerTrpTrpProGlyArgGlyLeuValSerTrpTrpMetTh 312  
Db 3278 GTGGGAGTGTCCCG-----CTGCCAGGTGCTGAGAGCTGTGTGTGTA 3325  
OY 312 rGlyArgSerArgAlaAlaGluGlyThrArgTrpValMetTrpPheGlyAspGlyLysPh 332  
Db 3326 CTAATTGCTTAGATTAAGCCAA-----GTTGTGTGTGCGCATATTAAGA 3370  
OY 332 eSerVal1aLysVal1GluLysLeu-----MetProLeuSerSerPheCysSerAlaPh 350  
3371 AGCAGTGTCTCAAGATGAGAAATGTCAAGATCTCAATGAGTCCCTCCAAAGCTTACC 3430  
OY 350 eHisGlnAlaThrTyraAsnLysGlnProMetTyraGlyLysAla1leTyraGluValLeuG 370  
Db 3431 TGCAGAAAGCAAGCTAAAGCTGTGAAA-AGAAAGAAAGAAAGCTTAAGACCAAGTAAAG 3489  
OY 370 nVal1aLysSerArgAla----- 376  
Db 3490 AAAGACAGCAAAAGAGAGAGCTGTGTAAGAAAGCTGTGACTCTAGTCAAAACCTACC 3549  
OY 377 -----GlyLysLeuPheProAlaCysHisAspSerAspGluSerAspSerGlyLys 393  
Db 3550 CCATCAGCAAGAGAGATCTGCTGCCCAAG-AAAAGCAGTAGAGAGCTCTCCACGAAA 3608  
OY 393 sAlaVal1GluVal1GlnAsnLysGlnMet1leGluTrpAlaLeuGlyGlyPheGlnProse 413  
Db 3609 GCCCGTCGAGAAAGAGAGAGAA-----GGGAATGTCTCGGCCCC 3650  
OY 413 rGlyProLysGlyLeuGln----- 419  
Db 3651 TGGGCTCGAATCCAAAGAGCCACACTCCAGCTTCCAGGAAGTCAAGCAAGAGTCTC 3710  
OY 420 -----ProProGluGluGlnLysAsn-----ProTyrlLysGlyVal 431  
Db 3711 CCAGCCAGCACTGTATCTCCGCTCCAGCACTTACTACAGACCGCCAAAGAAAGAGT 3770  
OY 431 1TyrlThrAspMetTrpVal1GluProGluAlaAla1aTyra1aProProProAlaLys 451

Db 3771 TCCCAAAACCACTCTTAGTAGAGCCCAAGAAAAG-----CAGCTTCCAGCACAGATC 3824  
OY 451 sLysProArgLysSer-----ThrThrG 459  
Db 3825 AGGTCCAGAGCAGAGCAAGAAAGAAAAGTGGCTCCCGCCCAAGTATCCCTGTAAACA 3884  
OY 459 uLysProLysVal1LysGlu----- 465  
Db 3885 AAAACCAAAAGAAAGAAAGAAACCACTCCGGTCAATAAAGCAGAGAAATGACAGCACTT 3944  
OY 466 -----1le1leAspGluArgThrArg----- 472  
Db 3945 GAAATCTCCAGCAGACTCTCTCAATAGGCAATAGTCTTAAGCAAAAATTCACAGATAG 4004  
OY 473 -----GluArgLeuValTyraGluValArgGlnLysCys-----ArgAsn1leGluAsp-- 488  
Db 4005 AGTCCACAGATCAGAGTGAAGCTTTAAGAGAGATTGTCAAGCAAAAATGTGCGAGAT 4064  
OY 489 -----LysAsnCysPheLeuGluCysAla1aTyraGlnTyraAspAspG 528  
Db 4167 TTGTAGACCTTCCACAAAGTTTGTAGAGAGAAAGAGCGCCCTCTGAGGAGCACACT 4226  
OY 528 yTyrlGlnSerTyraCysThr1leCys-----CysGly- 538  
Db 4227 GCAAAATGTTGTGTGTGTCTGCTGCAAAATCTGTACAGTTGTGGAAGCAATCAGAGC 4286  
OY 539 -GlyArgGluValLeuMetCysGlyLysAsnAsnCysCysArgCysPheCysVal1GluCy 558  
Db 4287 TACAAAGCAGCTGTGAGATG-----ATAAGTGCAGAAACAGCTATACCTGAGTG 4340  
OY 558 sValAspLeuLeuVal1GlyProGly---AlaAlaGlnAlaAla1leLysGlnAspProTr 577  
Db 4341 C-----CTGGACCAAACTAACCCACCAACCCACCAAGAAAGAAAGTGTG 4388  
OY 577 pAsnCyTyraMetCysGlyHisLysGly1aThrTyraGlyLeuLeuArgArgGluAspTr 597  
Db 4389 GATCTGTACCAAGTGTGTCTGCTGTAAGAGCTGTGATCCACAACTCCAGCAAAAGGTG 4448  
OY 597 pProSerArgLeuGlnMetPhePheAlaAsnHisAsp 610  
Db 4449 GGATGCAGAGTGTCTCATGATTTCTCATGTGTGATGAT 4488  
RESULT 10  
US-08-327-392-1  
; Sequence 1, Application US/08327392  
; Patent No. 5633136  
; GENERAL INFORMATION:  
; APPLICANT: Croce, Carlo  
; APPLICANT: Canaan, Eli  
; TITLE OF INVENTION: ALL-1 Polynucleotides and Monoclonal  
; TITLE OF INVENTION: Antibodies for Leukemia Detection and  
; TITLE OF INVENTION: Treatment  
; NUMBER OF SEQUENCES: 24  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock Washburn Kurtz MacLewicz & No. 5633136r1s  
; STREET: One Liberty Place - 46th Floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
; COMPUTER: IBM  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: WORDPERFECT 5.1





Db 2997 CTCTGTCCGAGGACCCCGGATTAAACATGCTCTGTCAGAACGACGAGCTGTGTGCGCTGGCGG 3056  
Qy 172 pGluSerLeuArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTy 192  
Db 3057 AAACGAGCTGTGTTCTCTGATGACATGCCACCCCTGAGT----- 3096  
Qy 192 rTyriLeSerLysArgLysArgAspGluTyrLeuAlaArgTrpLysArgGluAlaGluLy 212  
Db 3097 -----GCCTTACCATTG-----GAGA 3113  
Qy 212 sLysAlaLysValIleAlaValMetAsnAlaValGluGlnAsnGlnAlaSerGlyGluSe 232  
Db 3114 ACGAGAAAGATTTCTCTTCCATGGGAATGATGACAAGTCATCAATGTGTGGCTCA-- 3171  
Qy 232 rGlnLysValGluGlnAlaSerProProAlaValGlnGlnProThrAspProAlaSerPr 252  
Db 3172 -----GAAGATGCTGAACCTCTTGTCT-----CCACCCCATCAACCAATTAAC 3215  
Qy 252 oThrValAlaThrThrProGluProValGlyGlyAspAlaGlyAspLysAsnAlaThrLy 272  
Db 3216 TGTC-----ACTAGAAA 3227  
Qy 272 sAlaAlaAspAspGluProGluTyrGluAspGlyArgGlyPheGlyIleGlyGluLeuVa 292  
Db 3228 CAAGGCACCCAGGAACCTCCAGTAAAGAAAGACGTCG-ATC-----GAGGCGT 3277  
Qy 292 lTrpGlyLysLeuArgGlyPheSerTrpTrpProGlyArgIleValSerTrpTrpMetTh 312  
Db 3278 CTGGCAGTGTCCCG-----CTGCCAGGTGCTTGAGGAGTGTGTGTGTTGTA 3325  
Qy 312 rGlyArgSerArgAlaAlaGluGlyThrArgTrpValMetTrpPheGlyAspGlyLysPh 332  
Db 3326 CTAATGTCTAGATAAGCCCAA-----GTTTGTGTGTCGCAATATAAGA 3370  
Qy 332 eSerValValCysValGluLysLeu-----MetProLeuSerSerPheCysSerAlaPh 350  
Db 3371 AGCAGTGTGCAAGATGAGAAATGTCAGAAATCTACATGATGCTTCCAAAGCCTACC 3430  
Qy 350 eHisGlnAlaThrTyraAsnLysGlnProMetTyraGlyAlaAlaIleTyrGluValLeuGl 370  
Db 3431 TGCAAGACAGCTAAAGCTGTGAAAA-AGAAAGAGAAAAAAGTCTAAGACCAAGTGAAGAA 3489  
Qy 370 nValAlaSerSerArgAla----- 376  
Db 3490 AAAGACAGCAAGAGACGAGCTGTGTGAAGACGTGTGTGACCTCTAGTCAGAAACCTACC 3549  
Qy 377 -----GlyLysLeuPheProAlaCysHisAspSerAspGluSerAspSerGlyLy 393  
Db 3550 CCATCAGCAAGAGAGATCTGCCCCAAAG-AAAAGCAGTAGTGAGGCTCTCCACAGAAA 3608  
Qy 393 sAlaValGluValGlnAsnLysGlnMetIleGluTyrAlaLeuGlyGlyPheGlnProSe 413  
Db 3609 GCCCGTCCAGAAAAAGAGTGAAGAA-----GGGAATGTCTCGGCCCC 3650  
Qy 413 rGlyProLysGlyLeuGlu----- 419  
Db 3651 TGGGCTGAATCCCAACAGGCCACCACCTCCAGCTTCAGGAAGTCAAGCAGAGGTCTC 3710  
Qy 420 -----ProProGluGluGluLysAsn-----ProTyriLysGluVa 431  
Db 3711 CCAGCCAGCAGCTGTCTATCCCGCTCAGCCACCTACTACAGGACCGCCCAAGAAAGAGT 3770  
Qy 431 lTyThrAspMetTrpValGluProGluAlaAlaAlaTyrAlaProProProAlaLy 451  
Db 3771 TCCCAAAACCACTCTCTAGTGAGCCCCCAAGAAAAAG-----CAGCCTCCACCAAGATC 3824  
Qy 451 sLysProArgLysSer-----ThrThrGl 459  
Db 3825 AGTCCAGCAGCAGCAACAAAGAAAAAGTGGCTCCCGCCCAAGTATCCCTGTGTAACA 3884  
Qy 459 uLysProLysValLysGlu----- 465  
Db 3885 AAAACCAAGAAAGAAAGAAACCAACCTCCGGTCAATAAGCAGGAGAAATGCAGGCACTTT 3944

Qy 466 ----IleIleAspGluArgThrArg----- 472  
Db 3945 GAACATCTCTCAGCACTCTCTCCAATGGCAATAGTTCTTAAGCAAAAAATTTCCAGCAGATGG 4004  
Qy 473 ----GluArgLeuValTyrGluValArgGlnLysCys-----ArgAsnIleGluAsp-- 488  
Db 4005 AGTCCACAGATCAGAGTGGACTTTAAGGAGGATTTGAAGCAGAAAAATGTGGGAGAT 4064  
Qy 489 -----IleCysIleSerCy 493  
Db 4065 GGGAGGCTTAGGAATCTTGACTTCTGTCTTATACACCCAGGGTGGTTGCTTCTCTG 4124  
Qy 493 sGlySerLeuAsnValThrLeuGluHisProLeuPheIleGlyGlyMetCysGlnAsnCy 513  
Db 4125 TCCAGCTAGTGGG-----CATGTAGAGTTTGTG-----TATTGCCAAGTCTG 4166  
Qy 513 s-----LysAsnCysPheLeuGluCysAlaTyrGlnTyrAspAspGl 528  
Db 4167 TTGTGAGCCCTTCCACAAGTTTGTGTTAGAGGAGAAACGAGCGCCCTCTGGAGGACCAGCT 4226  
Qy 528 yTyriGlnSerTyriCysThrIleCys-----CysGly----- 538  
Db 4227 GGAAATTTGTGTGCTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 4286  
Qy 539 -GlyArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCy 558  
Db 4287 TACAAAGCAGCTCTGTGAGTGT-----ATAAGTGGCGAAACAGCTATCACCTGAGTG 4340  
Qy 558 sValAspLeuLeuValGlyProGly-----AlaAlaGlnAlaAlaIleLysGluAspProTr 577  
Db 4341 C-----CTGGACCAAACTACCCCAACCAACCAACCAACCAACCAACCAACCAACCA 4388  
Qy 577 pAsnCysTyriMetCysGlyHisLysGlyThrTyriGlyLeuLeuArgArgGluAspTr 597  
Db 4389 GATCTGTACCAAGTGTGCTGCTGTAAGAGCTGTGGATCCCAACTCCAGCGCAAGGGTG 4448  
Qy 597 pProSerArgLeuGlnMetPhePheAlaAsnAsnHisAsp 610  
Db 4449 GGATGCACAGCTGTCTCATGATTTCTCACTGTGTCTATGAT 4488

## RESULT 12

US-08-545-860D-1  
Sequence 1, Application US/08545860D  
Patent No. 6040140  
GENERAL INFORMATION:  
APPLICANT: Croce, Carlo  
TITLE OF INVENTION: Diagnostics, Therapeutics and Methods  
TITLE OF INVENTION: for Detection and Treatment of Acute Leukemias  
TITLE OF INVENTION: Resulting from Chromosome Abnormalities in the All-1 Region  
NUMBER OF SEQUENCES: 94  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz &  
ADDRESS: No. 604014018  
STREET: One Liberty Place, 46th floor  
CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/545,860D  
FILING DATE: 07-MAR-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/04496  
FILING DATE: 22-APR-1994  
PRIOR APPLICATION DATA:













```

QY 502 -----HisProLeu-----PheIleGlyMetCysGlnAsnCy 513
Db 1849 TCCAGCAGATGGAGTCCACAGCATCAGAGTGAGCATTTAAGTTGTGTATTCACCAAGTCTG 1908
QY 513 s-----LysAsnCySpheLeuGluCysAlaItyrGlnItyrAspAspG1 528
Db 1909 TTGTGAGCCCTTCACACAAGTTTGTGTAGGAGGAACGAGCCCTCTGAGGACAGCT 1968
QY 528 yTyrGlnSerTyrCysThrIleCys-----CysGly----- 538
Db 1969 GGAATAATGGTGTGTTCCTGCTGTTGCAAAATTCGTATCAGCTTTGTGGAAGGCACATCAGGC 2028
QY 539 -GlyArgGluValLeuMetCysGlyAsnAsnAsnCySAsyCysPheCysValGluCy 558
Db 2029 TACAAGCAGCTGCTGTGAGTGT-----AATAAGTCCGAAACAGATATCACCTTGAGTG 2082
QY 558 sValAspLeuValGlyProGly---AlaAlaGlnAlaAlaIleLysGluAspProTr 577
Db 2083 C-----CTGGGAGCCAAACTAACCCACCACCAAAACCACAAAGAAAGAAAGTCTG 2130
QY 577 pAsnCyStryrMetCysGlyHisIlyseGlyThrTyrGlyLeuLeuArgArgGluAspTr 597
Db 2131 GATCTGTACCAAGTGTCTTCCTGTGAAGAGCTGTGATCCACAACTCCAGCAGCAAGGCTG 2190
QY 597 pProSerArgLeuGlnMetPhePheAlaAsnAsnHisAsp 610
Db 2191 GGATGCACAGTGGTCTCATGATTTTCACACTGTGATGAT 2230

RESULT 15
US-08-465-713-4
; Sequence 4, Application US/08465713
; Patent No. 6121419
;
; GENERAL INFORMATION:
; APPLICANT: Rowley, Janet D.
; APPLICANT: Diaz, Manuel O.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: DETECTING GENE REARRANGEMENTS AND
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465,713
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/080,255
; FILING DATE: 17 JUNE 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:072/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 320-7200
; TELEFAX: (512) 474-7577
;
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; MOLECULE TYPE: DNA (genomic)
;
; US-08-465-713-4

```

Alignment Scores:	0.00427	Length:	4201
Pred. No.:	157.00	Matches:	148
Score:	34.47%	Conservative:	105
Percent Similarity:	20.16%	Mismatches:	274
Best Local Similarity:	3.20%	Indels:	211
Query Match:	3	Gaps:	32
DB:			
US-09-720-086-5 (1-908) x US-08-465-713-4 (1-4201)			
QY	10 SerSerSerSerLeuGluArgGluAspArgLysGluGluGluGluAsn	29	
DB	257 GCTGACAAAGCCTGGAGAGACCAAGATGAGAGAGACCCGGAGAGAAAAGAG	316	
QY	30 ArgGlyLysGluArgGluGlu-----ProSerAla	40	
DB	317 AATAAGCGGAGTCAGAGAAAGAAAAGGATCAGAAATTCAGAGTACTCT	376	
QY	41 ThrAlaArgLysValGlyArgProGlyArgLysArg---LysHisProProValGluSer	59	
DB	377 GCTTGATACCTGTGGGTAGGGTTTCCAAAGAGAGTGTGGTGAAGATCTGCCACT	436	
QY	60 SerAspThrProLysAspProAlaValThrThrLysSerGlnProMetAlaGlnAspSer	79	
DB	437 TCATCTTCGCGCAA-----AAAGCAACGGGGGGAAGAAAGCTTCATCATGATTTCT	490	
QY	80 GlyPro-----SerAspLeuLeuProAsnGlyAspLeuGluLys	92	
DB	491 GGGACTGATATTACTCTGTGACTCTGGGGATACACAGCTGT-CAAAACCAAATACT	549	
QY	93 ArgSerGluProGlnProGluGluGlySerProAlaAlaGlyGlnLysGlyAlaPro	112	
DB	550 TATAAGAAAGGAGAGAGAAATCTGSAAAAAAACCAACTTGGACCTCGGCCCACTGCCCC	609	
QY	113 AlaGluGlyGluGlyThrGluThrProPro-----GluAlaSerArgAlaValGlu	122	
DB	610 ATCCCTGGAGAGAGAAAACCTCTGCTTCCACTCTTCACTGACACTGTTAAACA	669	
QY	123 -----GluAlaSerArgAlaValGlu	129	
DB	670 TTCACCTTCCATAGAGCTCCATGTGCTGCTCAGGACAGACAAAGCTTCCAAAGACTGACAA	729	
QY	130 AsnGlyCysCysValThrLysGluGlyArgGlyAla-----	141	
DB	730 GAGGTTGCCAGCTCTTAATAAAGGCCAAAGCTCAGCTCTGCAAGTTGGAAGAGTAA	789	
QY	142 -----SerAlaGlyGluGlyLys-GluGlnLysGlnThr	152	
DB	790 GAGCTTTAAACAACCAAGCCAGCCCAAGACAGAGGTAAGAAAGTCACTCATCAGAGAC	849	
QY	152 AsnIleGluSerMetLysMetGluGlySerArgGlyArgLeuArgGlyGlyLeuGlyTr	172	
DB	850 CTCCTGCGCAGACCCCGGATTAACATGTCTGCAGAGAGCAGCTTGGCCCTTGGCCG	909	
QY	172 pGluSerSerLeuArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTyr	192	
DB	910 AAAACGAGCTGTCTTCTCGATGACATGCCACCTGAGT-----	949	
QY	192 rTyrIleSerLysArgLysArgAspGluTrpLeuAlaArgTrpLysArgGluAlaGlu	212	
DB	950 -----GCTTACCATGG-----GAGA	966	
QY	212 sLysAlaLysValIleAlaValMetAsnAlaValGluGluAsnGlnAlaSerGlyGluSe	232	
DB	967 ACAGAGAAAGATTTTGTCTTCATGGGGAATGATGACAAAGTCAATTCATGCTGCTCA--	1024	
QY	232 rGlnLysValGluGluAlaSerProProAlaValGlnGlnProThrAspProAlaSerPr	252	
DB	1025 -----GAAGATCTGAACCTTGT-----CCACCATCAACCAATTAAACC	1068	
QY	252 oThrValAlaThrThrProGluProValGlyGlyAspAlaGlyAspLysAsnAlaThrLys	272	

```
Db 1069 TCTC-----ACTAGAAA 1080
Qy 272 sAlaAlaAspGluProGluTyrGluAspGlyArgGlyPheGlyIleGlyGluLeuVa 292
Db 1081 CAAGGCACCCAGGACCTCCAGTAAGAAGAGCGTCG-ATC-----GAGCGCGT 1130
Qy 292 lTrpGlyLysLeuArgGlyPheSerTrpTrpProGlyArgIleValSerTrpTrpMetTh 312
Db 1131 GTGGCGAGTGCCTCGG-----CTGCCAGTGCCTCGAGGACTGTGGTGTGTTGA 1178
Qy 312 rGlyArgSerArgAlaAlaGluGlyThrArgTrpValMetTrpPheGlyAspGlyLysph 332
Db 1179 CTAATTGCTTAGATAAGCCCAA-----GTTTGGTGTGCTCCCAATATAAAGA 1223
Qy 332 eSerValValCysValGluLysLeu-----MetProLeuSerSerPheCysSerAl 349
Db 1224 AGCAGTGTCTGAAGATGAGAAATGTGCAATCTACTACAAATGGATGCTTCCAAAGCCT 1283
Qy 349 aPheHisGlnAlaThrTyrAsnLysGlnProMetTyrArgLysAlaIleTyrGluValle 369
Db 1284 ACCTGCAGAGCAAGCTAAGCTGTGAAGA-AGAAAGAGAAAGTCTAAGACCAAGTGA 1342
Qy 369 uGlnValAlaSerSerArgAla-----376
Db 1343 AAGAAAGACAGCAAGACGACGAGTGTGTGAAGAACGTGGTGGACTCTAGTCAGAAACCT 1402
Qy 377 -----GlyLysLeuPheProAlaCysHisAspSerAspGluSerAspSerGl 392
Db 1403 ACCCCATCAGCAAGAGGATCTCTGCCCCAAAG-AAAAGCAGTAGTAGGCTCTCCACG 1461
Qy 392 yLysAlaValGluValGlnAsnLysGlnMetIleGluTrpAlaLeuGlyGlyPheGlnPr 412
Db 1462 AAAAGCCGTGCGAGAAAGAGAGTGAAGAA-----GGCAATGTCTCGGC 1503
Qy 412 oSerGlyPro-----LysGlyLeuGluProProGluGluLysAsnProTyLysGl 430
Db 1504 CCTGGGCTGAATCCAAACACAGGCCACCACTCCAGTCTCCAGGAAGTCAAGCAAGCAGT 1563
Qy 430 uValTyrThrAspMetTrpValGluProGluAlaAlaTyrAlaProPro-----447
Db 1564 CTCCAGCCAGCAGCTGTCATCCCGCTCAG-----CCACCTACTACAGG 1608
Qy 448 -ProProAlaLysLysProArgLysSerThrThrGluLysProLysVallys-----464
Db 1609 ACCGCCAAGAAGAAGTTCCTCCAAACCACTCTCTAGTAGGCCCAAGAAAGACGCTCC 1668
Qy 465 -----GluIleAspGluArgThrArgGluArgLeuVal-----Ty 477
Db 1669 ACCACAGAAATCAGGTCCAGAGCAGAGCAACACAGAAAAAGTGGCTCCGCCCAAGTAT 1728
Qy 477 rGluValArgGlnLysCysArgAsnIleGlu-----AspIleCysIleSe 492
Db 1729 CCTGTAAAAACAAAACCAAAAGAAAGAAACCACCTCCGGTCAATAAGCAGAGAA 1788
Qy 492 rCysGlySerLeuAsnVal-----ThrLeuGlu-----501
Db 1789 TGCAGGCATTTGAACATCCTCAGCACTCTCTCCAATGGCAATAGTCTTAAGCAAAAAAT 1848
Qy 502 -----HisProLeu-----PheIleGlyMetCysGlnAsnCy 513
Db 1849 TCCAGCAGATGGAGTCCACAGGATCAGAGTGGACTTTAAGTTGTGTATTGCCAAGTCTG 1908
Qy 513 s-----LysAsnCysPheLeuGluCysAlaTyrGlnTyrAspAspGl 528
Db 1909 TTGTAGAGCCCTCCACAAAGTTTGTGTAGAGGAGACAGGCGCCCTCTGGAGGACCACT 1968
Qy 528 yTyrGlnSerTyrCysThrIleCys-----CysGly-----538
Db 1969 GGAAAAATTGGTGTGTGTCGTGTCGTAATTTCTGTACAGTTTGTGGAAGGCAACATCAGC 2028
Qy 539 -GlyArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCy 558
Db 2029 TACAAAGCAGCTGCTGGAGTGT-----AATAAGTCCGAAACACAGTATACCCCTGAGTG 2082
```

```
Qy 558 sValAspLeuLeuValGlyProGly---AlaAlaGlnAlaAlaIleLysGluAspProTr 577
Db 2083 C-----CTGGACCAAACTACCCCAACCAACCCACAAAGAGAGAAAGTCTG 2130
Qy 577 pAsnCysTyrMetCysGlyHisLysGlyThrTyrGlyLeuLeuArgArgGluAspTr 597
Db 2131 GATCTGTACCAAGTGTGTTCTGCTGTAGAGCTGTGGATCCCACTCCAGGCAAGGGTG 2190
Qy 597 pProSerArgLeuGlnMetPhePheAlaAsnAsnHisAsp 610
Db 2191 GGATGCACAGTGTGCTCATGATTTCTCACTGTGTATGAT 2230
```

Search completed: November 25, 2002, 01:45:03  
Job time : 312.239 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2002, 08:18:31 ; Search time 84.7727 seconds  
(without alignments)

15175.990 Million cell updates/sec

Title: US-09-720-086-2

Perfect score: 4195  
Sequence: 1 gaattccggcgccggcggtc.....aaaaaaaaaaaaaaaa 4195

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Sequences: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/prodata/1/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/prodata/1/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/prodata/1/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/prodata/1/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/prodata/1/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/prodata/1/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	548.4	13.1	2077	4	US-09-276-531-47 Sequence 47, Appl
2	49.4	1.2	7218	1	US-08-232-463-14 Sequence 14, Appl
3	45.2	1.1	7218	1	US-08-232-463-14 Sequence 14, Appl
4	43.2	1.0	1332	4	US-09-333-423-1 Sequence 1, Appl
5	42.8	1.0	812	4	US-09-091-097-7 Sequence 7, Appl
6	42.4	1.0	1046	1	US-08-361-467B-4 Sequence 4, Appl
7	42.4	1.0	1046	1	US-08-484-332C-4 Sequence 4, Appl
8	42.2	1.0	2239	4	US-09-196-390-1 Sequence 1, Appl
9	42	1.0	1378	4	US-09-149-476-208 Sequence 208, App
10	42	1.0	2323	4	US-09-149-476-24 Sequence 24, Appl
11	41.8	1.0	271	2	US-08-731-272A-29 Sequence 29, Appl
12	41.4	1.0	2065	4	US-09-370-473-5 Sequence 5, Appl
13	41.4	1.0	2091	4	US-09-813-818-1 Sequence 1, Appl
14	40.8	1.0	270	2	US-08-520-678A-30 Sequence 30, Appl
15	40.8	1.0	270	4	US-08-897-126-30 Sequence 30, Appl
16	40.8	1.0	790	4	US-09-461-474-13 Sequence 13, Appl
17	40.8	1.0	3637	1	US-08-445-640-3 Sequence 3, Appl
18	40.8	1.0	3637	3	US-08-170-558-3 Sequence 3, Appl
19	40.8	1.0	3637	3	US-08-447-314-3 Sequence 3, Appl
20	40.8	1.0	3637	3	US-08-445-461-3 Sequence 3, Appl
21	40.4	1.0	1578	4	US-09-416-050A-1 Sequence 1, Appl
22	40.4	1.0	1578	4	US-09-664-800-1 Sequence 1, Appl
23	40.4	1.0	1578	4	US-09-665-309-1 Sequence 1, Appl
24	40.4	1.0	1578	4	US-09-661-569-1 Sequence 1, Appl
25	40.2	1.0	635	1	US-08-455-633A-35 Sequence 35, Appl
26	40.2	1.0	635	1	US-08-416-336-5 Sequence 5, Appl
27	40.2	1.0	635	2	US-08-456-460C-35 Sequence 35, Appl

28	40.2	1.0	635	5	PCT-US94-05354-35	Sequence 35, Appl
29	40.2	1.0	1100	3	US-09-248-335-53	Sequence 53, Appl
30	40.2	1.0	1466	4	US-08-984-919A-10	Sequence 10, Appl
31	40.2	1.0	1466	4	US-08-984-919A-12	Sequence 12, Appl
32	40.2	1.0	1472	4	US-08-781-420-10	Sequence 10, Appl
33	40.2	1.0	1472	4	US-08-781-420-12	Sequence 12, Appl
34	40.2	1.0	1472	4	US-08-874-102-10	Sequence 10, Appl
35	40.2	1.0	1472	4	US-08-874-102-12	Sequence 12, Appl
36	40.2	1.0	1472	4	US-09-006-595A-10	Sequence 10, Appl
37	40.2	1.0	1472	4	US-09-006-595A-12	Sequence 12, Appl
38	40.2	1.0	1493	6	5340934-5	Patent No. 5340934
39	40.2	1.0	1875	4	US-08-984-919A-46	Sequence 46, Appl
40	40.2	1.0	1875	4	US-08-984-919A-48	Sequence 48, Appl
41	40.2	1.0	1881	4	US-08-874-102-46	Sequence 46, Appl
42	40.2	1.0	1881	4	US-08-874-102-48	Sequence 48, Appl
43	40	1.0	2852	3	US-09-027-137-2	Sequence 2, Appl
44	40	1.0	2852	4	US-09-344-441-2	Sequence 2, Appl
45	39.8	0.9	337	2	US-09-032-684-11	Sequence 11, Appl

#### ALIGNMENTS

RESULT 1  
US-09-276-531-47  
Sequence 47, Application US/09276531  
Patent No. 6183968  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Yue, Henry  
APPLICANT: Reddy, Roopa  
APPLICANT: Guegler, Karl J.  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
NUMBER OF SEQUENCES: 134  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/276,531  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/079,677  
FILING DATE: March 27, 1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Lynn E. Murry, Ph.D.  
REGISTRATION NUMBER: 42,918  
REFERENCE/DOCKET NUMBER: PA-0008 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2077 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: TESTTUT02



[illegible]

RESULT 3  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 570067

? GENERAL INFORMATION:  
 ? APPLICANT: DONNER, F.  
 ? APPLICANT: SCHEIFELINGER, F.  
 ? APPLICANT: FALKNER, F. G.  
 ? TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
 ? NUMBER OF SEQUENCES: 52  
 ? CORRESPONDENCE ADDRESS:  
 ? ADDRESSEE: Foley & Lardner  
 ? STREET: 1800 Diagonal Road, Suite 500  
 ? CITY: Alexandria

? COUNTRY: USA  
 ? ZIP: 22313-0299  
 ? COMPUTER READABLE FORM:  
 ? MEDIUM TYPE: Floppy disk  
 ? COMPUTER: IBM PC compatible  
 ? OPERATING SYSTEM: PC-DOS/MS-DOS  
 ? SOFTWARE: Patent in Release #1.0, Version #1.25  
 ? CURRENT APPLICATION DATA:  
 ? APPLICATION NUMBER: US/08/232,463  
 ? FILING DATE:

CLASSIFICATION 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935.313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 INMUL  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-9300

```

; INFORMATION FOR SEQ ID NO: 14:
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 7218 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;

```

;  
; IMMEDIATE SOURCE:  
; CLONE: PTZgpt-Fls  
US-08-232-463-14

Query Match	1.1%;	Score 45.2;	DB 1;	Length 7218;
Best Local Similarity	4.3%;	Pred. No. 0.035;		
Matches	17;	Conservative 212;	Mismatches 165;	Indels 0;
			Gaps	0;

QY 2769 AGCTGCTGGCAGGTCCTGSAGTGTACCCGTTATCAGACTGTGGCCCCCTGAAG 2828

QY 2829 ACTACTTGCCTGTGAATAGTTCTTACCACGACTGGGAGCTTCGGTCAAGGCCAGTGC 2888

```
Db      1096 yyyyyyyyyyyyyyyyyyyyyyy 1155
```

```

Db 1156 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1215
      : : : : : : : : : : : : : : : : : : : : : : : : : :
      : : : : : : : : : : : : : : : : : : : : : : : : : :

```

[illegible]

3009 TGCAGGGGAGCCCCCGTGCTCTCCCTCCGTGTCACAGCTCAGACCTGGCTGCTTAGAGTAG 3068

[illegible]

Db 1336 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1395

```

0Y      3129 AIGCCTTCTTTACCCCTCGAGTTATCACCTC 3162
          :   :::::  :::::  :::  :::
Db      1396 XXXXXXXXXXXXXXXXXXXXXXXXXXXXX 1429

```

RESULT 4  
US-09-333-423-1

Patent No. 6265636  
GENERAL INFORMATION:  
APPLICANT: Randall, Douglas  
APPLICANT: Thelen, Jay  
APPLICANT: Miernyk, Jan  
APPLICANT: Muszynski, Michael  
APPLICANT: Sewalt, Vincent  
TITLE OF INVENTION: Pyruvate Dehydrogenase Kinase  
FILE OF INVENTION: Polynucleotides, Polypeptides and Uses thereof  
FILE REFERENCE: 0818

EARLIER APPLICATION NUMBER: 60/089,998  
EARLIER FILING DATE: 1998-06-19

```
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
; INCHW1 1333
```

```

; TYPE: DNA
; ORGANISM: Zea mays
;

```

```

; NAME/KEY: CDS
; LOCATION: (55) ... (1095)
;
TTC-00 232 423 1

```

Query Match	1.0%	Score 43.2	DB 4	Length 1332
PostgreSQL similitude	0.0%	Score 0.0	DB 0	Length 0.0

Accession	Sequence	Position
QY 3128	GATGGCTTCTTTTACCCCTCCGAGTTTATCACTCAGAAGTGATGGCTAAGATACCAAA	3187
Db 1236	GATTGTGATCGTAATCTCTAATCGTTTATGACTGATTGTACTGTCCAAAAA	1295



```

: Patent No. 5767374
: GENERAL INFORMATION:
: APPLICANT: De Greef, Willy
: APPLICANT: Van Emmele, John
: APPLICANT: De Oliveira, Dulce B.
: APPLICANT: De Souza, Maria-Helena
: APPLICANT: Van Montagu, Marc
: TITLE OF INVENTION: PLANTS WITH MODIFIED FLOWERS, SEEDS OR
: TITLE OF INVENTION: EMBRYOS
: NUMBER OF SEQUENCES: 13
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
: STREET: P.O. Box 1404
: CITY: Alexandria
: STATE: Virginia
: COUNTRY: United States
: ZIP: 22313-1404
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/484,332C
: FILING DATE: 7-JUNE-1995
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/361,467
: FILING DATE: 22-DEC-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/681,492
: FILING DATE: 04-APR-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/EP90/01275
: FILING DATE: 01-AUG-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: EP 89 402 224.3
: FILING DATE: 04-AUG-1989
: ATTORNEY/AGENT INFORMATION:
: NAME: Schulman, Robert M.
: REGISTRATION NUMBER: 31,196
: REFERENCE/DOCKET NUMBER: 010830-093
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 836-6620
: TELEFAX: (703) 836-2021
: ORAMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1046 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: IMMEDIATE SOURCE:
: CLONE: 3C9
: US-08-484-332C-4

Query Match 1.0%; Score 42.4; DB 1; Length 1046;
Best Local Similarity 55.4%; Pred. No. 0.069;
Matches 82; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```

```

RESULT 8
US-09-196-390-1
: Sequence 1, Application US/09196390
: Patent No. 6307125
: GENERAL INFORMATION:
: APPLICANT: Block, Martina
: APPLICANT: Loitz, Horst
: APPLICANT: Luticke, Stephanie
: APPLICANT: Walter, Lennart
: APPLICANT: Froberg, Claus
: APPLICANT: Kossmann, Jens
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING ENZYMES
: TITLE OF INVENTION: FROM WHEAT WHICH ARE INVOLVED IN STARCH
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: James F. Haley, Jr., c/o Fish & Neave
: STREET: 1251 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: United States of America
: ZIP: 10020
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/196,390
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DE 196 21 588.9
: FILING DATE: 29-MAY-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: DE 196 36 917.7
: FILING DATE: 11-SEP-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/EP97/02793
: FILING DATE: 28-MAY-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Haley, Jr., James F.
: REGISTRATION NUMBER: 27,794
: REFERENCE/DOCKET NUMBER: AGREVO-9
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 596-9000
: TELEFAX: (212) 596-9090
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2239 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to mRNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Triticum aestivum L.
: STRAIN: cv. Florida
: HAPLOTYPE: ca. 21 d Caryopses
: IMMEDIATE SOURCE:
: LIBRARY: cDNA library in pBluescript sk (-)
: CLONE: Tasss
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 3..2017
: US-09-196-390-1

Query Match 1.0%; Score 42.2; DB 4; Length 2239;
Best Local Similarity 54.1%; Pred. No. 0.12;
Matches 86; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```

QY 3073 GCATGTCGTCATGTTCTTACCTGAACTTTAAACTTGAAGTAGTAGATGAGTGC 3132

Mon Nov 25 08:18:10 2002

us-09-720-086-2.rni

Db	2042	GCCTTCTGTGTA	TGTTGTCCTT	TAGTCGACAA	TAATTTGACCTGTTGGAGATTTT	2101
Oy	3133	CTTTCTTTTACCTCCT	CGAGTTTACATCAGAACTGATGGCTAGATACCA	AAAAAACA		3192
Db	2102	ATCTTGCTGCTGTTTT	TTTTTTTAATCAAAAGAGGGGTTCT	CCGATTTCA	TAAAAA	2161
Oy	3193	AACAAAAACAGAA	CAAAAAACAAAAAACCTCAACA			3231
Db	2162	AAAAA	AAAAA	AAAAA	AAAAA	2200

RESULT 9  
US-09-149-476-208  
; Sequence 208, Application US/09149476  
; Patent No. 6420526  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 186 Human Secreted proteins

```

FILE REFERENCE: PZ002PI
CURRENT APPLICATION NUMBER: US/09/149,476
CURRENT FILING DATE: 1998-09-08
EARLIER APPLICATION NUMBER: PCT/US98/04493
EARLIER FILING DATE: 1998-03-06
EARLIER APPLICATION NUMBER: 60/040,162
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,333
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/038,621
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,626
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,334
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,336
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,163
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/047,600
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,615
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,597
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,502
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,633
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,583
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,617
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,618
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,503
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,592
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,581
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,584
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,500
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,587
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,492
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,598
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,613
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,582

```

1	EARLIER	FILING DATE:	1997-05-23
2	EARLIER	APPLICATION NUMBER:	60/047,596
3	EARLIER	FILING DATE:	1997-05-23
4	EARLIER	APPLICATION NUMBER:	60/047,612
5	EARLIER	FILING DATE:	1997-05-23
6	EARLIER	APPLICATION NUMBER:	60/047,632
7	EARLIER	FILING DATE:	1997-05-23
8	EARLIER	APPLICATION NUMBER:	60/047,601
9	EARLIER	FILING DATE:	1997-05-23
10	EARLIER	APPLICATION NUMBER:	60/043,580
11	EARLIER	FILING DATE:	1997-04-11
12	EARLIER	APPLICATION NUMBER:	60/043,568
13	EARLIER	FILING DATE:	1997-04-11
14	EARLIER	APPLICATION NUMBER:	60/043,314
15	EARLIER	FILING DATE:	1997-04-11
16	EARLIER	APPLICATION NUMBER:	60/043,569
17	EARLIER	FILING DATE:	1997-04-11
18	EARLIER	APPLICATION NUMBER:	60/043,311
19	EARLIER	FILING DATE:	1997-04-11
20	EARLIER	APPLICATION NUMBER:	60/043,671
21	EARLIER	FILING DATE:	1997-04-11
22	EARLIER	APPLICATION NUMBER:	60/043,674
23	EARLIER	FILING DATE:	1997-04-11
24	EARLIER	APPLICATION NUMBER:	60/043,669
25	EARLIER	FILING DATE:	1997-04-11
26	EARLIER	APPLICATION NUMBER:	60/043,312
27	EARLIER	FILING DATE:	1997-04-11
28	EARLIER	APPLICATION NUMBER:	60/043,313
29	EARLIER	FILING DATE:	1997-04-11
30	EARLIER	APPLICATION NUMBER:	60/043,672
31	EARLIER	FILING DATE:	1997-04-11
32	EARLIER	APPLICATION NUMBER:	60/043,315
33	EARLIER	FILING DATE:	1997-04-11
34	EARLIER	APPLICATION NUMBER:	60/048,974
35	EARLIER	FILING DATE:	1997-06-06
36	EARLIER	APPLICATION NUMBER:	60/056,886
37	EARLIER	FILING DATE:	1997-08-22
38	EARLIER	APPLICATION NUMBER:	60/056,877
39	EARLIER	FILING DATE:	1997-08-22
40	EARLIER	APPLICATION NUMBER:	60/056,889
41	EARLIER	FILING DATE:	1997-08-22
42	EARLIER	APPLICATION NUMBER:	60/056,893
43	EARLIER	FILING DATE:	1997-08-22
44	EARLIER	APPLICATION NUMBER:	60/056,630
45	EARLIER	FILING DATE:	1997-08-22
46	EARLIER	APPLICATION NUMBER:	60/056,878
47	EARLIER	FILING DATE:	1997-08-22
48	EARLIER	APPLICATION NUMBER:	60/056,662
49	EARLIER	FILING DATE:	1997-08-22
50	EARLIER	APPLICATION NUMBER:	60/056,872
51	EARLIER	FILING DATE:	1997-08-22
52	EARLIER	APPLICATION NUMBER:	60/056,882
53	EARLIER	FILING DATE:	1997-08-22
54	EARLIER	APPLICATION NUMBER:	60/056,637
55	EARLIER	FILING DATE:	1997-08-22
56	EARLIER	APPLICATION NUMBER:	60/056,903
57	EARLIER	FILING DATE:	1997-08-22
58	EARLIER	APPLICATION NUMBER:	60/056,888
59	EARLIER	FILING DATE:	1997-08-22
60	EARLIER	APPLICATION NUMBER:	60/056,879
61	EARLIER	FILING DATE:	1997-08-22
62	EARLIER	APPLICATION NUMBER:	60/056,880
63	EARLIER	FILING DATE:	1997-08-22
64	EARLIER	APPLICATION NUMBER:	60/056,894
65	EARLIER	FILING DATE:	1997-08-22
66	EARLIER	APPLICATION NUMBER:	60/056,911
67	EARLIER	FILING DATE:	1997-08-22
68	EARLIER	APPLICATION NUMBER:	60/056,636
69	EARLIER	FILING DATE:	1997-08-22
70	EARLIER	APPLICATION NUMBER:	60/056,874
71	EARLIER	FILING DATE:	1997-08-22
72	EARLIER	APPLICATION NUMBER:	60/056,910
73	EARLIER	FILING DATE:	1997-08-22



```

1 EARLIER APPLICATION NUMBER: 60/056, 864
2 EARLIER FILING DATE: 1997-08-22
3 EARLIER APPLICATION NUMBER: 60/056, 631
4 EARLIER FILING DATE: 1997-08-22
5 EARLIER APPLICATION NUMBER: 60/056, 845
6 EARLIER FILING DATE: 1997-08-22
7 EARLIER APPLICATION NUMBER: 60/056, 892
8 EARLIER FILING DATE: 1997-08-22
9 EARLIER APPLICATION NUMBER: 60/057, 761
10 EARLIER FILING DATE: 1997-08-22
11 EARLIER APPLICATION NUMBER: 60/047, 595
12 EARLIER FILING DATE: 1997-05-23
13 EARLIER APPLICATION NUMBER: 60/047, 599
14 EARLIER FILING DATE: 1997-05-23
15 EARLIER APPLICATION NUMBER: 60/047, 588
16 EARLIER FILING DATE: 1997-05-23
17 EARLIER APPLICATION NUMBER: 60/047, 585
18 EARLIER FILING DATE: 1997-05-23
19 EARLIER APPLICATION NUMBER: 60/047, 586
20 EARLIER FILING DATE: 1997-05-23
21 EARLIER APPLICATION NUMBER: 60/047, 590
22 EARLIER FILING DATE: 1997-05-23
23 EARLIER APPLICATION NUMBER: 60/047, 594
24 EARLIER FILING DATE: 1997-05-23
25 EARLIER APPLICATION NUMBER: 60/047, 589
26 EARLIER FILING DATE: 1997-05-23
27 EARLIER APPLICATION NUMBER: 60/047, 593
28 EARLIER FILING DATE: 1997-05-23
29 EARLIER APPLICATION NUMBER: 60/047, 614
30 EARLIER FILING DATE: 1997-05-23
31 EARLIER APPLICATION NUMBER: 60/043, 578
32 EARLIER FILING DATE: 1997-04-11
33 EARLIER APPLICATION NUMBER: 60/043, 576
34 EARLIER FILING DATE: 1997-04-11
35 EARLIER APPLICATION NUMBER: 60/047, 501
36 EARLIER FILING DATE: 1997-05-23
37 EARLIER APPLICATION NUMBER: 60/043, 670
38 EARLIER FILING DATE: 1997-04-11
39 EARLIER APPLICATION NUMBER: 60/056, 632
40 EARLIER FILING DATE: 1997-08-22
41 EARLIER APPLICATION NUMBER: 60/056, 664
42 EARLIER FILING DATE: 1997-08-22
43 EARLIER APPLICATION NUMBER: 60/056, 876
44 EARLIER FILING DATE: 1997-08-22
45 EARLIER APPLICATION NUMBER: 60/056, 881
46 EARLIER FILING DATE: 1997-08-22
47 EARLIER APPLICATION NUMBER: 60/056, 909
48 EARLIER FILING DATE: 1997-08-22
49 EARLIER APPLICATION NUMBER: 60/056, 875
50 EARLIER FILING DATE: 1997-08-22
51 EARLIER APPLICATION NUMBER: 60/056, 862
52 EARLIER FILING DATE: 1997-08-22
53 EARLIER APPLICATION NUMBER: 60/056, 887
54 EARLIER FILING DATE: 1997-08-22
55 EARLIER APPLICATION NUMBER: 60/056, 908
56 EARLIER FILING DATE: 1997-08-22
57 EARLIER APPLICATION NUMBER: 60/048, 964
58 EARLIER FILING DATE: 1997-06-06
59 EARLIER APPLICATION NUMBER: 60/057, 650
60 EARLIER FILING DATE: 1997-09-05
61 EARLIER APPLICATION NUMBER: 60/056, 884
62 EARLIER FILING DATE: 1997-08-22
63 EARLIER APPLICATION NUMBER: 60/057, 669
64 EARLIER FILING DATE: 1997-09-05
65 EARLIER APPLICATION NUMBER: 60/049, 610
66 EARLIER FILING DATE: 1997-06-13
67 EARLIER APPLICATION NUMBER: 60/061, 060
68 EARLIER FILING DATE: 1997-10-02

```

Query Match	1.0%	Score 42;	DB 4;	Length 1378;
Best Local Similarity	60.5%	Pred. No. 0.11;		
Matches 69;	Conservative	0;	Mismatches 45;	Indels 0;
			Gaps	0

QY	3118	TAGGTAGTAAAGATGGCTTTCTTTTACCCCTCGATTCATCACTCAGAAGGATGGCTAA	3177
Db	1248	TATGTACAACTGATGCTTGTCTTATTTTATATAATTATCAGAGTGAAAAA	1307
QY	3178	GATACCAAAAAACAAACAAAAACAGAAACAAAAACAAAAAAACCTCAACA	3231
Db	1308	AA	1361

```

RESULT 10
US-09-149-476-24
; Sequence 24, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; EARLIER FILING DATE: 1998-03-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23

```

EARLIER APPLICATION NUMBER: 60/047,612  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,632  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,601  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,580  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,568  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,314  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,569  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,311  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,671  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,674  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,569  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,312  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,313  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,672  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,315  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/048,974  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/056,886  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,877  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,889  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,893  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,630  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,878  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,662  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,872  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,882  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,637  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,903  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,888  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,879  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,880  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,894  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,911  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,636  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,874  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,910  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,864  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,631

EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,845  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,892  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057,761  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/047,595  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,599  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,588  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,585  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,586  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,590  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,594  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,589  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,593  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,614  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,578  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,576  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/047,501  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,670  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/056,632  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,664  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,876  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,881  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,909  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,875  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,862  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,887  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,908  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/048,964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/057,650  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/056,884  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/057,669  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/049,610  
EARLIER FILING DATE: 1997-06-13  
EARLIER APPLICATION NUMBER: 60/061,060  
EARLIER FILING DATE: 1997-10-02

Query Match 1.0%; Score 42; DB 4; Length 2323;  
Best Local Similarity 60.5%; Pred. No. 0.14;  
Matches 69; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

OY 3118 TAGGTAGTAAAGTGGCTTTCTTTTACCTCTGAGTTTATCACTCAGAGTGTGCTAA 3177  
DB 2186 TATGTACAACTGATGCTTGTCTTTTATTAATAATTTATCAGAGTGAATAAAAAA 2245



SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/897.126  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/520.678  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Henderson, Melodie W.  
REGISTRATION NUMBER: 37,848  
REFERENCE/DOCKET NUMBER: 6029-6836  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 314-727-5188  
TELEFAX: 314-727-6032  
TELEX:  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 270 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-897-126-30

	Query Match	1.0%	Score 40.8	DB 4	Length 270
	Best Local Similarity	67.9%	Pred. No. 0.088	27	Indels 0
	Matches 57	Conservative	0	Mismatches	0
					Gaps 0
3176	AAAGATACCCAAAAAACAACCAAAAAACAGAAACAAAAAACCCTCAACGCTC	3235			
QY					
DB					
171	AAGGAC	112			
QY					
DB					
3236	TCCTTAGTACTCAGGTTCATGCTGC	3259			
QY					
DB					
111	CCTAAGGTCGGAGTGTTTAAGCTGC	88			
QY					
DB					

Search completed: November 22, 2002, 09:19:58  
Job time : 212.773 secs

	Query Match	1.0%;	Score 40.8;	DB 2;	Length 270;
	Best Local Similarity	67.9%;	Pred. NO. 0.088;		
	Matches 57;	Conservative 0;	Mismatches 27;	Indels 0;	Gaps 0;
3176	AAAGATACCAAAAAAAAAACAAACAAAAACAGAAACAAAAACAAAAAACTTCAACAGCTC	3235			
171	AAGAGGAC	112			
3236	TCTTAGTACTCAGGTTTCATGCTGC	3259			
111	CCTAAGGTCGGAGTGTTAAGCTGC	88			

```

RESULT 15
US-08-897-126-30/c
; Sequence 30, Application US/08897126
; Patent No. 6297003
; GENERAL INFORMATION:
; APPLICANT: Rice, Charles M.
; APPLICANT: Kolykhalov, Alexander A.
; TITLE OF INVENTION: NOVEL 3' TERMINAL SEQUENCE OF HEPATITIS
; TITLE OF INVENTION: C VIRUS GENOME AND DIAGNOSTIC AND THERAPEUTIC USES THEREOF
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howell & Hafertkamp, L.C.
; STREET: 7733 Forsyth Blvd., Suite 1400
; CITY: St. Louis
; STATE: MO
; COUNTRY: USA
; ZIP: 63105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 24, 2002, 12:10:37 ; Search time 67.8633 Seconds  
(without alignments)  
3854.741 Million cell updates/sec

Title: US-09-720-086-8  
Perfect score: 4566  
Sequence: 1 MKGDRHLNGEDAGGRS.....MSVEVIRHLFAPLKOYFACE 853

## Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Command line parameters:

MODEL=frame+ p2n.model -DEV=xlh  
O=/cg2\_1/USPTO.spool/US09720086/runat.18112002.092957.29457/app.query.fasta.1.4252  
-DB=Issued\_Patents\_NA -QPM=faetap -SUFFIX=rni -MINMATCH=0.1 -DOFCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09720086 @CGN.1.1.108 @runat.18112002.092957.29457 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

## Database :

Issued\_Patents\_NA.\*  
1: /cg2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cg2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cg2\_6/ptodata/1/ina/5A.COMB.seq:\*  
4: /cg2\_6/ptodata/1/ina/5B.COMB.seq:\*  
5: /cg2\_6/ptodata/1/ina/PCTUS.COMB.seq:\*  
6: /cg2\_6/ptodata/1/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	903	19.8	2077	4	US-09-276-531-47 Sequence 47, Appl
2	148.5	3.3	6328	4	US-08-913-832A-1 Sequence 1, Appl
3	148.5	3.3	6328	4	US-08-913-832A-1 Sequence 1, Appl
4	146	3.2	2355	4	US-08-913-159-12 Sequence 12, Appl
5	144.5	3.2	3522	4	US-09-484-9708-8 Sequence 8, Appl
6	137.5	3.0	4481	4	US-09-041-886-18 Sequence 18, Appl
7	137.5	3.0	36778	4	US-09-105-537-5 Sequence 5, Appl
8	137.5	3.0	38506	3	US-09-320-878-19 Sequence 19, Appl
9	133	2.9	35524	3	US-08-923-137-1 Sequence 1, Appl
10	130.5	2.9	2376	2	US-08-760-745-4 Sequence 4, Appl
11	129.5	2.8	80161	3	US-09-036-987A-1 Sequence 1, Appl
12	129.5	2.8	80161	4	US-09-370-700-1 Sequence 1, Appl

13	126	2.8	8580	4	US-09-491-772-1 Sequence 1, Appl
14	125.5	2.7	2874	4	US-09-179-558-54 Sequence 54, Appl
15	125.5	2.7	17612	3	US-08-911-853-29 Sequence 29, Appl
16	125.5	2.7	17612	4	US-09-479-409-29 Sequence 29, Appl
17	125.5	2.7	17612	4	US-09-479-453-29 Sequence 29, Appl
18	124.5	2.7	4689	4	US-09-105-537-34 Sequence 34, Appl
19	124.5	2.7	4235	4	US-09-199-637A-1 Sequence 1, Appl
20	124	2.7	36778	4	US-09-105-537-5 Sequence 5, Appl
21	124	2.7	38506	3	US-09-320-878-19 Sequence 19, Appl
22	123	2.7	11220	4	US-09-105-537-32 Sequence 32, Appl
23	122	2.7	2151	4	US-09-627-216A-9 Sequence 9, Appl
24	122	2.7	2439	4	US-09-624-693A-18 Sequence 18, Appl
25	122	2.7	5392	2	US-08-403-852D-1 Sequence 1, Appl
26	122	2.7	5392	3	US-08-510-646B-1 Sequence 1, Appl
27	122	2.7	5392	4	US-09-231-818-1 Sequence 1, Appl
28	121.5	2.7	13842	4	US-09-105-537-30 Sequence 30, Appl
29	121.5	2.7	30001	1	US-08-125-468-1 Sequence 1, Appl
30	121.5	2.7	30001	2	US-08-474-933-1 Sequence 1, Appl
31	121	2.7	1170	3	US-09-307-621-1 Sequence 1, Appl
32	120.5	2.6	3059	4	US-09-179-558-61 Sequence 1, Appl
33	120	2.6	43280	2	US-08-804-227C-1 Sequence 1, Appl
34	119	2.6	2151	4	US-09-627-216A-7 Sequence 7, Appl
35	118.5	2.6	16836	4	US-09-147-236-1 Sequence 1, Appl
36	118.5	2.6	16836	4	US-09-147-236-1 Sequence 1, Appl
37	118.5	2.6	4403765	4	US-09-103-840A-2 Sequence 2, Appl
38	118.5	2.6	4411529	4	US-09-103-840A-1 Sequence 1, Appl
39	118	2.6	7295	2	US-08-487-826B-15 Sequence 15, Appl
40	117.5	2.6	3688	6	5248670-4 Patent No. 5248670
41	117.5	2.6	34063	4	US-09-453-702B-96 Sequence 96, Appl
42	117.5	2.6	4411529	4	US-09-103-840A-1 Sequence 1, Appl
43	117	2.6	8982	3	US-08-976-255-5 Sequence 5, Appl
44	116.5	2.6	2623	2	US-08-973-675-1 Sequence 1, Appl
45	116.5	2.6	11219	1	US-07-642-734C-1 Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-09-276-531-47  
Sequence 47, Application US/09276531  
Patent No. 6183968  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Yue, Henry  
APPLICANT: Reddy, Roopa  
APPLICANT: Guegler, Karl J.  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
NUMBER OF SEQUENCES: 134  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/276,531  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/079,677  
FILING DATE: March 27, 1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

```
; NAME: Lynn E. Murry, Ph.D.
; REGISTRATION NUMBER: 42,918
; REFERENCE/DOCKET NUMBER: PA-0008 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 955-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 47:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2077 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: TESTTUT02
; CLONE: 1271435
; US-09-276-531-47
```

```
Alignment Scores:
Pred. No.: 3 45e-76 Length: 2077
Score: 803.00 Matches: 183
Percent Similarity: 89.32% Conservative: 1
Best Local Similarity: 89.83% Mismatches: 21
Query Match: 19.78% Indels: 4
DB: 4 Gaps: 1
```

US-09-720-086-8 (1-853) x US-09-276-531-47 (1-2077)

```
QY 648 ThrGlyArgLeuPheGluPheTyrHisLeuLeuAsnTyrSerArgProLysGluGly 667
DB 3 GCGGCNANTGCAACGATCTNNR-AATGTGATCA-GCCAGGAAAGGCTGTATGAGGAC 60
QY 668 ThrGlyArgLeuPheGluPheTyrHisLeuLeuAsnTyrSerArgProLysGluGly 687
DB 61 AGG---CGGCTCTTCGAAATTTTC-CACCTGTGTAATTTACTCACGCCCAAGAGGGT 116
QY 688 AspAspArgProPheTyrMetPheGluAsnValAlaMetLysValGlyAspLys 707
DB 117 GATGACCGNCGTCTCTTGGATGTTTGAGATGTTGTAGNCTCGTCCGATTCCGCCAGA 176
QY 708 ArgAspIleSerArgPheLeuGluCysAsnProValMetIleAspAlaIleLysValSer 727
DB 177 GGGGACATCTACGGTCTCGGAGTGTAAATCCAGTGATGATTTGATGCCATCAAGTTTCT 236
QY 728 AlaAlaHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMetAsnArgProValIle 747
DB 237 GCTGTCTACAGGCGCGGATACTTCTGGGGCAACCTACCCGGGATGAACAGGCCCGTGATA 296
QY 748 AlaSerLysAsnAspLysLeuGluLeuAsnProGlyMetAsnArgIleAlaLys 767
DB 297 GCATCAAGAAATGATAAATCTCGNGCTGCAGGACTGCTTGAATACAAATAGGATAGCCCAAG 356
QY 768 LeuLysLysValGlnThrIleThrThrLysSerAsnSerIleLysGlnGlyLysAsnGln 787
DB 357 TTTAAAGAAAGTACAGACAATTAACCAAGTCGAACTCGATCAAAACAGGGGAAAAACCAA 416
QY 788 LeuPheProValValMetAsnGlyLysGluAspValLeuTyrCysThrGluLeuGluArg 807
DB 417 CTTTTCCCTGTTGTCTAATGATGCAAGAGATGTTTNGTGTGCACTGAGCTCGAAAGG 476
QY 808 IlePheGlyPheProValHisTyrThrAspValSerAsnMetGlyArgGlyAlaArgGln 827
DB 477 NTCTTTGGCTTTCTGTGCACTACAGACGTGTCTCAACATGGGCGGTGGTGGCCGCCAG 536
QY 828 LysLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAlaProLeuLys 847
DB 537 AAGCTGCTGGGAAGGTCTCTGGAGCGTGCCTGTATCCGACACCTCTTCGCCCTCTGAAG 596
QY 848 AspTyrPheAlaCysGlu 853
DB 597 GACTACTTTGCATGTGAA 614
RESULT 2
US-08-913-832A-1
```

```
; Sequence 1, Application US/08913832A
; Patent No. 6329517
; GENERAL INFORMATION:
; APPLICANT: Seelig, Hans Peter
; TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN
; FILE REFERENCE: 8484-0030-999
; CURRENT APPLICATION NUMBER: US/08/913.832A
; CURRENT FILING DATE: 1998-01-12
; PRIOR APPLICATION NUMBER: PCT/DE96/00444
; PRIOR FILING DATE: 1996-03-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 6328
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(5736)
US-08-913-832A-1
Alignment Scores:
Pred. No.: 0.00229 Length: 6328
Score: 148.50 Matches: 108
Percent Similarity: 34.57% Conservative: 87
Best Local Similarity: 19.15% Mismatches: 193
Query Match: 3.25% Indels: 176
DB: 4 Gaps: 26
US-09-720-086-8 (1-853) x US-08-913-832A-1 (1-6328)
QY 14 AlaGlyGlyArgGluAspSerIle-----LeuValAsnGlyAlaCys----- 27
DB 37 GCGGGCAGTGAGGAGGAGATATGATGCACATTTTGAACAACAGCGCTGCCCCACCCAC 96
QY 28 SerAspGlnSerSerAspSerProIleLeuGluAlaIleArgThrProGluIleArg 47
DB 97 CCAGAAATGAAGAGGACCCAGAGAGGATTTGTTCAGAAACAGACACTCCAAAGCTCAAG 156
QY 48 GlyArgSerSerSerArg-----LeuSerLysArgGluVal 60
DB 157 AAGAAAGAAAGAGCTTAAGAAACCTCGGGACCTAAATCCCTAAGAGCAAGCGCAAAA 216
QY 61 SerSerLeuLeuSerTyrThrGlnAspLeuThrGlyAspGlyAspGlyGluAspGlyAsp 80
DB 217 AAGGAGCGTATGCTCTTTATCGCGCAGCTG--GGGGACAGCTCTGGGGAGGGG---- 267
QY 81 GlySerAspThrProValMetProLysLeuPheArgGluThrArgThrArgSerGluSer 100
DB 268 -----CCAGAGTTTGGAGAGGAGGAGGAGGAGGCTCTCGCTCGCTCAGACAGT 315
QY 101 ProAlaValArgThrArgAsnAsnSerValSerSerArgGluArgHisArgProSer 120
DB 316 GAGGGC-----AGCGACTATACTCTCTGCAAGAGAGAGAGAGAGAGCTTGA 363
QY 121 ProArgSerThrArgGlyArgGlnGlyArgAsnHisValAspGluSerProValGluPhe 140
DB 364 CTTAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 411
QY 141 ProAlaThrArgSerLeuArgArgAlaThrAlaSerAlaGlyThrProTrpProSer 160
DB 411 -----CAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 411
QY 161 ProProSerSerTyrLeuThrIleAspLeuThrAspAspThrGluAspThrHisGlyThr 180
DB 412 -----CAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 441
QY 181 ProGlnSerSerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyMet 200
DB 442 CCTAAATCATCT-----GCTCAGCTCTCTGGAAGAC-----TGGGGCATG 480
QY 201 GluSerProGlnValGluAlaAspSerClyAspGlyAsp-----SerSerGlu 216
```

Dd 481 GAA-----GACATTGACCAGTGTCTCAGAGGAGAT 513

Qy 217 TycIAspBgLYLysgluPheGLYllegLYAspLeuValTrpGLYylsleYsgLYphe 236  
|||:::  
Dd 514 TATGAACCCCTCACCACTAC-----AAGCCTTC 543

Qy 237 SerTrpTrpProAlaMetValSerTrpLys-----AlathrSerLysArg 252  
|||:::  
Dd 544 AGCCACTTTGTGCAGACCCCTCATTTGGTCCAAAAATCCCAAGATTGCTGTCCAAAGATG 603

Qy 253 GluAlaMetSerGlyMetArgTrpValGluTrpPheGlyAspGLYlysPhe----- 269  
|||:::  
Dd 604 ATGATGCTTTGGTGTCAGAAAATGGCGGAGTTCCAGTACCATAAACCCCTTCAAAGCACT 663

Qy 270 -----SerGluValSerAlaAspLYsLeuValAlaleGlyLeuPheSerGlnHisPhe 287  
|||:::  
Dd 664 TCTGGGGCATCTAGTGGCAGCTGCCGGAGCAGCAGCGGTAGCTGTGGTGAG----- 714

Qy 288 AsnLeuAlaThrPheAsnLeuLYsLeuValSerTYrArgLYsAlaMetYrHisAlalaGlu 307  
|||:::  
Dd 715 AGCATGTGCACACCCACTAGGTTGCACACCACTCCCCCTGTGGAGGTCCATTATCCC 774

Qy 308 LysAlaArgValArgAlaGlyLYsThrPheProSerSerProGlyAspSerLeuGluAsp 327  
|||:::  
Dd 775 AAGGCCAACAACCAAGAGGGCAAGGT-CCCAATGT- 810

Qy 328 GlnLeuLYsPrometLeuGluTrpAlaHisGlyglYpHeLysProthrGlyllegGly 347  
|||:::  
Dd 811 CGGAGCAACCCC----- 822

Qy 348 LeuLYsProAsnAsnThrGlnProValAlaAsnLYsSerLYsValArgArgAlaGlySer 367  
|||:::  
Dd 823 ---AAGGGCAGCCCTCGTGTACTGTATGCCAAGAAGCCTTAACCCAAGAAAGTAGCTCCC 879

Qy 368 ArgLYsLeuGluSerArgLYserGlyTrgLYsAsnLYsThrArgArgArgThralAlaAspAspSer 387  
|||:::  
Dd 880 CTGAANAATCAAGCTGGAGGTTTGGTTCCAAAGCGTAAGAGATCTCGAAGTGAG----- 933

Qy 388 AlaThrSerAspTYrCYsProAlaParPolyArgLeuLYsThrAsnCYsTYrAsnAsnGly 407  
|||:::  
Dd 933 ----- 933

Qy 408 LysAspArgGlyAspGluAspGlnSerArgGluGlnMetAlaSerAspValAlaAsnAsn 427  
|||:::  
Dd 934 -----GATGATGACTTAGTAGTGTGAATCTGACTTCGATGTGCGCAGATTCAT 981

Qy 428 LysSerSerLeuGluAspGlyCYsLeu-----SerCYsGlyArgLYsAsnProVal 444  
|||:::  
Dd 982 AGCTATTCTGTTCTGTGATGTTCACACGACCGTAGTAGCGCAGCCCAAGAAACTCCGA 1041

Qy 445 SerPheHisProLeuPheGluGlyLYsLeuCYsGlnThrCYsArgAspArgPheLeuGlu 464  
|||:::  
Dd 1042 ACCACTAAAAGAAAAAGAAAGGCGAGGAGGAGTGACTGCTGTGATGCT----- 1092

Qy 465 LeuPheTYrMetLYrAspAspAspGlyTYrGlnSerTYrCYsThrValCYsCYsGluGly 484  
|||:::  
Dd 1093 -----FATGACACAGAC---CACGAGACTATTGTCAGAGTGTGCAGCAAGGC 1137

Qy 485 ArgGluLeuLeuLeuCYsSerAsnThrSerCYsCYsArgCYsPheCYsValGluCYsLeu 504  
|||:::  
Dd 1138 GGTTGAGATCATCTCTGTGTGAT-----ACCTGTCCCCCGTTCACACATGTGTCTGCTG 1191

Qy 505 GluValLeuValGlyThrGlyThralAlaGluAlaLYsLeuGlnGluProTrpSerCYs 524  
|||:::  
Dd 1192 GAT-----CCCGACATGGAGAAAGGCTCCGAGGGCAAG-----TGAAGCTGC 1233

Qy 525 TYrMetCYsLeuProGln-ArgCYsHisGlyValLeuArgArg-----Ar 539

Dd 1234 CCACACTGCGAAGAAAGAGCATCCAGTGGGAAGCTAAAGAGACAATTGGAGGGGTGAG 1293

Qy 539 gLYsAspTrp 542  
|||:::

```

DB      1294 GAGATCCTCG 1303

RESULT 3
US-09-249-181A-1
; Sequence 1, Application US/09249181A
; Patent No. 6440679
; GENERAL INFORMATION:
; APPLICANT: Seelig, Hans Peter
; APPLICANT: Renz, Manfred
; TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN
; FILE REFERENCE: 8484-0059-999
; CURRENT APPLICATION NUMBER: US/09/249,181A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 08/913,832
; PRIOR FILING DATE: 1998-01-12
; PRIOR APPLICATION NUMBER: PCT/DE96/00444
; PRIOR FILING DATE: 1996-03-08
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 6328
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(5736)
US-09-249-181A-1

Alignment Scores:
Pred. No.:      0.00229      Length:      6328
Score:          148.50      Matches:      108
Percent Similarity: 34.57%      Conservative: 87
Best Local Similarity: 19.15%      Mismatches: 193
Query Match:      3.25%      Indels:      176
DB:                4      Gaps:      26

US-09-720-086-8 (1-853) x US-09-249-181A-1 (1-6328)

QY      14  AAlaGlyArgGluAspSerIle-----IeuValAsnGlyValAcys----- 27
DB      37  GCGGGCACTAGAGAGAGAGATGATGATGACCTTTTGGACAAACAGCCCTGCCACCCAC 96
QY      28  SerAspGlnSerSerAspSerProProlIleuGluAlaIleArgThrProGluIleArg 47
DB      97  CCAGAAAAATGAAGAGAGACCAGAAAGAGGATTGTCAAAAAACAGAGACTCCAAAGCTCAAG 156
QY      48  GlyArgArgSerSerSerArg-----LeuSerLysArgGluVal 60
DB      157  AAGAGAGAAAAAGCCTTAAGAAACCTCGGAGACCCTAAATCCTTAAGACCAAGCCCAAAA 216
QY      61  SerSerLeuLeuSerTyrThrGlnAspLeuThrGlyAspGlyAspGlyValAsp 80
DB      217  AAGAGCGCTATGCTCTTATGCGGCGACGCTG---GGGAGACAGCTCTGGGAGGGG----- 267
QY      81  GlySerAspThrProValMetProLysLeuAspLeuArgGluThrArgThrArgSerGluSer 100
DB      268  -----CCAAGATTGTGGAGAGAGAGAAAGAGGTGCTCGCGGCTCAACACAG 315
QY      101  ProAlaValArgThrArgAsnAsnAsnSerValSerSerArgGluAlaGlnHisArgProSer 120
DB      316  GAGGGC-----ACCGACTATATCTCTGGCAGAAAGAAAGAAAGAAAGCTTGG 363
QY      121  ProArgSerThrArgGlyArgGlnGlyArgAsnHisValAspGluSerProValGluPhe 140
DB      364  CCTTAAGAAAGAGAGAAAGCAATCCAAAGCGGAAAGAGAGAGAGAGAG- 411
QY      141  ProAlaThrArgSerLeuArgArgAlaThrAlaSerAlaGlyThrProTrpProSer 166
DB      411  ----- 411
QY      161  ProProSerSerTyrIleuThrIleAspLeuThrAspAspThrGluAspThrHisGlyThr 180
DB      412  -----GAGGATGATGATGATGATGATTCAAAGAG 441

```



```
QY 181 ProGlnSerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyMet 200
|||:|||||
Db 442 CCTAAATCATCT-----GCTCAGCTCCTGGAAGAC-----TGGGGCATG 480
QY 201 GluSerProGlnValGluAlaAspSerGlyAspGlyAsp-----SerSerGlu 216
|||:|||||
Db 481 GAA-----GACATTGACACGCTGTTCTCAGAGGAGAT 513
QY 217 TyrGlnAspGlyGlyGluPheGlyIleGlyAspLeuValTrpGlyLysIleGlyPhe 236
|||:|||||
Db 514 TATCGAACCTCACCACATAC-----AAGCCCTTC 543
QY 237 SerTrpTrpProAlaMetValValSerTrpLys-----AlaThrSerLysArg 252
|||:|||||
Db 544 AGCCAGTTTGTGACAGCCCTCATCTGTCGCAAAATCCCAAGATTGTTCTTCAAGATG 603
QY 253 GlnAlaMetSerGlyMetArgTrpValGlnTrpPheGlyAspGlyLysPhe-----269
|||:|||||
Db 604 ATGATGGTGTGGGTGCAAAATGGCGGGAGTTCCAGTACCATAACCCCTTCAAGGCAGT 663
QY 270 -----SerGluValSerAlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPhe 287
|||:|||||
Db 664 TCTGGGCATCAGTGCAGCTCGGCAGCAGCAGCGGTAGCTGTGGTGGAG-----714
QY 288 AsnLeuAlaThrPheAsnLysLeuValSerTyrArgLysAlaMetTyrHisAlaLeuGlu 307
|||:|||||
Db 715 AGCATGGTGACAGCCACCTGAGGTGACACCACTCCCTCCCTGTGGAGGTGCCTATCCGC 774
QY 308 LysAlaArgValAlaGlyLysThrPheProSerSerProGlyAspSerLeuGluAsp 327
|||:|||||
Db 775 AAGCCAAAGACCAAGGAGGCAAGGT---CCCAATGCT-----810
QY 328 GlnLeuLysProMetLeuGluTrpAlaHisGlyGlyPheLysProThrGlyIleGluGly 347
|||:|||||
Db 811 CGGAGGAAGCCC-----822
QY 348 LeuLysProAsnAsnThrGlnProValValAsnLysSerLysValArgAlaGlySer 367
|||:|||||
Db 823 ---AAGGCAGCCCTCGTGTACTGTATGCCAAGAACCTTAACCCCAAGAAAGTAGTCC 879
QY 368 ArgLysLeuGluSerArgLysTyrGluAsnLysThrArgArgThrAlaAspSer 387
|||:|||||
Db 880 CTGAAATCAAGCTGGAGGTTTGGTTCACAGCGTAAGAGATCCTCGAGTGAG-----933
QY 388 AlaThrSerAspTyrCysProAlaProLysArgLeuLysThrAsnCysTyrAsnAsnGly 407
|||:|||||
Db 933 -----933
QY 408 LysAspArgGlyAspGlnSerArgGluGlnMetAlaSerAspValAlaAsnAsn 427
|||:|||||
Db 934 -----GATGATGACTTAGATGTGAATCTGACTTCGATGATGCCAGTATCAAT 981
QY 428 LysSerSerLeuGluAspGlyCysLeu-----SerCysGlyArgLysAsnProVal 444
|||:|||||
Db 982 AGCTATTCTGTTCTGTGTTGTTCCACCCAGCTAGTAGCCGACGCGCAAAACTCGGA 1041
QY 445 SerPheHisProLeuPheGluGlyLeuCysGlnThrCysArgAspArgPheLeuGlu 464
|||:|||||
Db 1042 ACCACTAAAGAAAGAAAGCGGAGGAGGAGTCTGCTGTGGTGTG-----1092
QY 465 LeuPheTyrMetTyrAspAspGlyTyrGlnSerTyrCysThrValCysCysGluGly 484
|||:|||||
Db 1093 -----TATGACACAGAC---CACCAGGACTATTGCCAGGTGTCCCAAGGC 1137
QY 485 ArgGluLeuLeuLeuSerAsnThrSerCysCysArgCysPheCysValGluCysLeu 504
|||:|||||
Db 1138 GGTGAGATCATCTGTGTGAT-----ACCTGTCCCGGTGCTTACCACATGGTGTGCTG 1191
QY 505 GluValLeuValGlyThrGlyThrAlaAlaGluAlaLysLeuGlnGluProTyrPheSerCys 524
|||:|||||
Db 1192 GAT-----CCGACATGAGAGGCTCCGAGGGCAAG-----TGGAGGTGC 1233
```

```
QY 525 TyrMetCysLeuProGln-ArgCysHisGlyValLeuArgArg-----Ar 539
|||:|||||
Db 1234 CCACACTGGCAGAGGAAGGCATCCAGTGGAGAGCTAAAGAGGACAATTCCGGGGTGAG 1293
QY 539 gLysAspTyr 542
|||:|||||
Db 1294 GAGATCCTGG 1303

RESULT 4
US-08-913-159-12
; Sequence 12, Application US/08913159
; Patent No. 6300109
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Plasmid-derived type II
; TITLE OF INVENTION: restriction-modification systems from Lactococcus lactis
; NUMBER OF SEQUENCES: 14
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,159
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DK 0179/95
; FILING DATE: 17-FEB-1995
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2355 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Lactococcus lactis subsp. cremoris
; STRAIN: W39
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 744..1283
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start=744
; OTHER INFORMATION: /product="LlaDII restriction endonuclease"
; OTHER INFORMATION: /evidence=EXPERIMENTAL
; OTHER INFORMATION: /gene="ORF"
; OTHER INFORMATION: /number=1
; OTHER INFORMATION: /standard_name="Gene coding for R.LlaDII"
; OTHER INFORMATION: /label="R-LlaDII"
; OTHER INFORMATION: /note="The first ten amino acids in this sequence may be
; OTHER INFORMATION: doubtful. However, from base 773 this reading frame gives a h
; OTHER INFORMATION: homology with the Bsp6I endonuclease"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1392..2342
; IDENTIFICATION METHOD: experimental
; OTHER INFORMATION: /codon_start=1392
; OTHER INFORMATION: /product="LlaDII methylase"
; OTHER INFORMATION: /evidence=EXPERIMENTAL
; OTHER INFORMATION: /gene="ORF"
; OTHER INFORMATION: /number=2
; OTHER INFORMATION: /standard_name="Gene coding for M.LlaDII"
; OTHER INFORMATION: /label="m-LlaDII"
; OTHER INFORMATION: /note="The sequence shows 60 % identity and 76 % similarity
; OTHER INFORMATION: with the Bsp6I methylase."
US-08-913-159-12

Alignment Scores:
Pred. No.: 0.000753 Length: 2355
Score: 146.00 Matches: 74
Percent Similarity: 36.39% Conservative: 49
```

Best Local Similarity: 21.89% Mismatches: 125  
Query Match: 3.20% Indels: 90  
DB: 4 Gaps: 17

US-09-720-086-8 (1-853) x US-08-913-159-12 (1-2355)

```
OY 575 ILeArgValLeuSerLeuPheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeu 594
DB 1395 TTGAAATGCTCTCTTTTCGCCGAGTTGGCGAATTGATTGTTTGAATAATGCA 1454
OY 595 GLyIleLysValGlyLysTyrValAlaSerGluValCysGluGluSerIleAlaValGly 614
DB 1455 GGTTTCAAAACA-----ATATATGCTAATGAATTT-----GATTAATGCTGCTGAT 1502
OY 615 ThrValLysHisGluGluValAsnIleLysTyrValAspAspValArgAsnIleThrLysLys 634
DB 1503 ACTTTTGAATGAACCTTGACGCTTAAGGTA-----GACCGACGCGATTAATGATGTA 1556
OY 635 AsnIleGluGluThrGlyProPheAspLeuValIleGlyLysProCysAsnAspLeu 654
DB 1557 CAACCTGATGAATACACAGATTGATATGATGACAGGTTTCTTCCCAAGCCTTT 1616
OY 655 SerAsnValAsnProAlaArgLysGlyLeu-----TyrGluGlyThrGlyArgLeuPhe 672
DB 1617 TCT--ATTGCTGATTATCGTCAAGGCTTTAAGCATGAACAAGGTCGAGGTAATCTTTT 1673
OY 673 PheGluPheTyrHisLeuLeuAsnTyrSerArgProLysGluGlyAspAspArgProPhe 692
DB 1674 TTGAACTTGTTCGATTATTAGAAACAAACAAACCTCGTGTGCA----- 1718
OY 693 PheTyrMetPheGluAsnValValAlaMetLysValGlyAspLysArgAspIleSerArg 712
DB 1719 -----TCTTTGAAATGTTAAATCTGTTCTCAGATAGCGGGAACACATTTGA 1772
OY 713 PheLeuGluCysAsnProVal-----MetIleAspAlaIleLys 725
DB 1773 GTTATTT--TGTTCTGAGTAGAAAGACTAGGATCAAGTATCTTTTCAAGTCTTAAT 1829
OY 726 ValSerAlaAlaHisArgAlaArgTyrPheTyrPglYasLeuProGlyMetAsnArgPro 745
DB 1830 GCTTCTGAA-----TATGAAATATATACCT--CAAAATAGAGNA 1865
OY 746 -----ValIleAspSerLysAsnAsp----- 752
DB 1866 CGTATCTATATTGTTGCTTTCAAAAATAAAAAGATTATGCAAAATTTGAACTACCAAA 1925
OY 753 -----LysLeuGluLeuGluAspCysLeuGluTyrAsnArg----- 764
DB 1926 TCTATACCTTTTAAACCAACGATTCAGATGTTATGTTTTCTTAAAAAACAAGACGAT 1985
OY 764 ----- 764
DB 1986 AAGTTCTACTATACCTCTGAAAGAAATMAATTTTGTATGATGATTAAGAAATATGACT 2045
OY 765 -----IleAlaLysLeuLysLysValGlnThrIleThrLysSerAsnSer 780
DB 2046 AATCAACGACATCATATACAGTGGCGTAGAGTTATGTAGAGAAACAAAGAAATTTTA 2105
OY 781 IleLysGln-----GlyLysAsnGlnLeuPheProVal 792
DB 2106 GTACCAACACTAACGCGTAAATATGGGACAGGTGGGCAATAT-----GTGCTATATATC 2159
OY 793 MetAsnGlyLysGluAspValLeuTyrCysThrGluLeuGlu-----ArgIlePheGly 810
DB 2160 CTTACATATAGCGAGATATTCTGTAATTAACACCAAGAAATGCTTTAAGCTTCAAGGT 2219
OY 811 PheProValHisTyrThrAspValSerAsnMetGlyArgGlyAlaArgGlnLysLeuLeu 830
DB 2220 TTCCCAAAAGAAATAT--AACTTCCAAACCAAAAGTAATGGAGATTAATATAACCAAGCA 2276
OY 831 GlyArgSerThrPheValProValIleArgHisLeuPheAlaProLeuLysAsp 848
DB 2277 GGAACAGTGTGTAGTACCAATTATAGAAAGAAATTCGAAAAAATCTTGACAT 2330
```

## RESULT 5

US-09-484-970B-8  
Sequence 8, Application US/09484970B  
Patent No. 6426186

## GENERAL INFORMATION:

APPLICANT: Jones, Karen A.  
APPLICANT: Volkmut, Wayne  
APPLICANT: Walker, Michael G.  
TITLE OF INVENTION: BONE REMODELING GENES  
FILE REFERENCE: PB-0014 US  
CURRENT APPLICATION NUMBER: US/09/484, 970B  
CURRENT FILING DATE: 2000-01-18  
NUMBER OF SEQ. ID NOS: 172  
SOFTWARE: PERL Program  
SEQ ID NO 8  
LENGTH: 3522

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc feature

OTHER INFORMATION: Incyte ID No. 6426186 249096.3

NAME/KEY: unsure

LOCATION: 656, 658, 663

OTHER INFORMATION: a, t, c, g, or other

US-09-484-970B-8

## Alignment Scores:

Pred. No.:	0.00206	Length:	3522
Score:	144.50	Matches:	135
Percent Similarity:	34.78%	Conservative:	81
Best Local Similarity:	21.74%	Mismatches:	233
Query Match:	3.16%	Indels:	173
DB:	4	Gaps:	32

US-09-720-086-8 (1-853) x US-09-484-970B-8 (1-3522)

```
OY 28 SerAspGlnSerSerAspSerProProlLeuGluAlaIleArgThrProGluIleArg 47
DB 766 ACAGACCAATCCACCAAGACCTTTTCAACAATAAGATTCCACGAAACCACTGACTA-- 822
OY 48 GlyArgSerSerSerArgLeuSerLysArgGluValSerSerLeuLysTyrThr 67
DB 823 -----GCAAGACCAACTCAGCGCCACACAGATTTTATCTGTAAGGCCCAAGACA 876
OY 68 GlnAspLeuThrGlyAspGlyAspGlyGluAspGlySerAspThrProValMet 87
DB 877 -----TCTGACAGCCACACATC 894
OY 88 ProlLysLeuPheArgGluThrArgThrArgSerGluSerProAlaValArg--ThrArg 106
DB 895 AGACCTGTTCTGAATAGACAACTACAGACCTTACAGGCCCAAAACCAAGTGGATGCC 954
OY 107 AsnAsnSerSerValSerSerArgGluArgHisArgProSerProArgSerThrArgGly 126
DB 955 AGTGGAAATGAGGTGGCAAGAGGGGTCAAGCAAGCAACCCAGGCCA-----TAGGT 1005
OY 127 ArgGlnGlyArgAsnHisValAspGluSer--ProValGluPheProAlaThrArg-- 144
DB 1006 GCTGATACAAATGTATCAGTGCATCTTACCCACCCCACTTAAAGACCGGAGCTCGCCGC 1065
OY 145 ---SerLeuArgArgArgAlaThrAlaSerAlaGlyThrProTyrPro-- 159
DB 1066 CCACCTTGCCACCCAGACTTACACACCCACGAAGAAACCTTTACCAACCAATATATGTC 1125
OY 160 -----SerProProSerSerTyrLeuThrIleAspLeuThrAspAsp--ThrGluAspTh 177
DB 1126 ACTGAAAGCCGGAAGTGCAGAAATCTTTCATCAGGGCCCATTAATACACACCCCT 1185
OY 177 rHisGlyThrProGlnSerSerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGly 197
DB 1186 GAGGTCAACACCCAGGCTTACTGTAACCTCTTGAGAGATATAGACAGATATATAAGCA 1245
```

QY 197 nGlyGlyMetGluSerProGlnValGluAlaAsePserGlyAseP-----G1 212  
Db 1246 A-----CCACAGTTTCTCTGCC-----TCTGGAGAAGAACTGGAAATATAAC 1287  
QY 212 yAspSerSer-GluTyrGlnAsePglyLysGluPheGlyIleGly-----A 227  
Db 1288 TGACTTTAGCTCAAGCCCAAGCAAGAACTGATCTCTTGGGAAGCCAAAGATTCAAAGG 1347  
QY 227 spLeuValTrpGlyLysIleLysGlyPheSerTrpTrpProAlaMetValSerTrp- 246  
Db 1348 ACCTCATGTGGATACATCCAAAGCCTGACAAAGCTGACAAAGCTCTGCTCCATTACTGACTCTGT 1407  
QY 247 --LysAlaThrSerLysArgGlnAlaMetSerGlyMet----- 258  
Db 1408 CAACGGTTTCCCAAGAGAGAGCCACAGAGGGGAATGCCACAGCCACCACAGAACCC 1467  
QY 259 -----ArgTrpValGlnTrpPheGlyAsePglyLysPheSerGluValSerA 274  
Db 1468 ACCCACCACCTCACTGTGTGTCACCGTGAAGGTGCCCTTCTCATTTGTGCTTGACT 1527  
QY 274 laAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPheAenLeuAlaThrPheAenL 294  
Db 1528 GGGAAAGCCCACTAAAT-----GACACTGTCACTGAATATG 1563  
QY 294 yLeuValSerTyrArgLysAlaMetTyrHisAlaLeuGluLysAlaArgValArgAlaG 314  
Db 1564 AAGTTATATCCAGAAATAATGGTCAATTCAGTGGGAAGAACAGTCCATTCAATGACAA 1623  
QY 314 lYlsThrPheProSer-----SerProGlyAsePserLeuGluAsePglneul 330  
Db 1624 ATCAGACATTTTCCACAGTAGAAATCTCGAAACCAACAGAGTTATGAATTCAGGTGA 1683  
QY 330 yProMetLeuGluTrpAlaHisGlyGlyPheLysProThr-----GlyIleGluG 347  
Db 1684 AACCCAAAACCGGTTGGTGAAGCCCGTTCAGTGGGAAGAACAGTCCAGTCACTGAAT 1743  
QY 347 lYleuLysProAenAenThrGlnProVal----- 356  
Db 1744 CAGCGGACCCCAAGAGTGAAGTGCAGCCAGTTTCTGCAGGAAGAGATGCCATCTGGACTGAAA 1803  
QY 357 -----ValAenLysSerLysValArgAlaG 366  
Db 1804 GACCTTTTAATTCACACTCTTACTCAGAGTGTAAGGCCAAACAATATGTCAAAAGGACAT 1863  
QY 366 lYserArgLysLeuGluSerArgLysTyrGluAenLysThrArgArgThr-----AlaA 385  
Db 1864 GGTATAAAAAATTTGTAGAGTGCAGCTGTGCACACTCTCTCAGATACAGATTTACTTGA 1923  
QY 385 spAspSerAlaThrSerAspTyrCysProAlaProLysArgLeuLysThrAenCystYrA 405  
Db 1924 GCGACTCCCTCACAGGAAATTT-----TATA 1950  
QY 405 spAenGlyLysAspArgGlyAseP-----G 413  
Db 1951 ACATAGGTGATCAGAGGGCCCATGGAGAAGATCACTGCCAGTTTGGGATTCATTTTAG 2010  
QY 413 luAsePglneulSerArgGluGlnMetAlaSerAsePValAlaAenAenLysSerLeuGluA 433  
Db 2011 ATGGACGCACCTGGGCAACACTCACTTCTGAC-----CAGTTACCAATCAAAAG 2058  
QY 433 spGlyCysLeuSerCysGlyArgLysAsePProValSerPheHisProLeuPheGluGlyG 453  
Db 2059 AAGGTTATTTTTCAGACAGTTCGCCAGGAACCTGTCCAAATTTGAGAAATA-----GGTG 2112  
QY 453 lYleuCysGlnThrCysArgAsePArgPheLeuGluLeuPheTyrMetTyrAsePAspG 473  
Db 2113 GTCACACCCAAATC-----AATTATGTTTCAGTGGTATGA-ATGTGGGAC----- 2155  
QY 473 lYTrGlnSerTyrCysThrValCysCysGluGlyArgGluLeuLeuLeu-----CysS 491  
Db 2156 --TACAATTCCTGGAAATGGTAGATGCTGCACAAAGTTACCTTCTGTTTCATCATTTGCA 2213  
QY 491 eAenThrSer-----CysCysArg-----CysPheCysValGluCysL 504

Db 2214 AACAAAAATCATTTGAAAAATACTATGCGCATTCATTTAAAGCTATTTTGTACTATGTA 2273  
QY 504 euGluVal-----LeuValGlyThrGlyThr-----AlaAlaG 515  
Db 2274 TAAAGTCTACAATCTAAATTAATAGCAATACTAGATGTTTATTATAGAAAAAGATTGCTG 2333  
QY 515 luAla-----LysLeuGlnGluPro----- 521  
Db 2334 AGAGTATTTATCATAGTGTTTTACAAAGTCCATTTTAAAGCAAGACTGATGTTTAAACAG 2393  
QY 522 -----TrpSerCysTyrMetCysLeuProGlnAArgCysHisGlyValLeuArg 537  
Db 2394 AATAACATTTTGGGAAGCTGGTCCCTATT-----CATGGTATTTTAAAGA 2440  
RESULT 6  
US-09-041-886-18  
; Sequence 18, Application US/09041886  
; Patent No. 6235872  
; GENERAL INFORMATION:  
; APPLICANT: Bredesen, Dale E.  
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence  
; TITLE OF INVENTION: Polypeptides and Methods of Use  
; NUMBER OF SEQUENCES: 72  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: United States  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION NUMBER: US/09/041,886  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LJ 2626  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4481 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 163..4099  
US-09-041-886-18  
Alignment Scores:  
Pred. No.: 0.0144 Length: 4481  
Score: 137.50 Matches: 108  
Percent Similarity: 33.46% Conservative: 63  
Best Local Similarity: 21.14% Mismatches: 211  
Query Match: 3.01% Indels: 132  
DB: 4 Gaps: 18  
US-09-720-086-8 (1-853) x US-09-041-886-18 (1-4481)  
QY 11 GluGluAsePAlaGlyGlyArgGluAsePserIleLeuValAenGlyAlaCysSerAsePglneul 30  
Db 1375 GAAGAAATATGTT-----GTAGTGTCTAGTATGATAGCAGTTTA 1416

```
OY 31 SerSerAspSerProProIleuGluAlaIleArgThrProGluIleArgGlyArg 50
|||
|||
|||
Db 1417 TCTTCGTACAGTGGCC---TTAGAAAAGATAACTCAGAAATTTTAAACGGGAA 1473
OY 51 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuSerTyrThrGlnAspLeu 70
|||
|||
|||
Db 1474 GCAGAGGCAACACAGTTAGCAGAAAGAAATTGAGTCAAGTCCAGTACAAAGCTCGAGTG 1533
OY 71 ThrGlyAspGlyAspGlyGluAspGlyAspGlySerAspThrProValMetProLysLeu 90
|||
|||
|||
Db 1534 GCCCTGGAAATATGATGATAGAGTAGAGAGAAAGAAATATACACAGCAGTT--- 1581
OY 91 PheArgGluThrArgThrArgSerGlySerProAlaValArgThrArgAsnAsnSer 110
|||
|||
|||
Db 1582 ---CAGAGAAATTCAGTAGAAGCGAGGAGCAGCATAAACACTAGAGAAATTAATAT 1638
OY 111 ValSerSerArgGluThrArgProSerProArgSerThrArgGlyValArgGlnGlyArg 130
|||
|||
|||
Db 1639 ATTCTCTCTGCAAGAAAGAAATAGAGAAAGTCAATCTCTGGAGAAAGTGGAGACGAATTC 1698
OY 131 AsnHisValAspGluSerProValGluPheProAlaThrArgSerLeuArgArgAla 150
|||
|||
|||
Db 1699 CCGCGTATGGCCAG---CCTGATCGGCTCCATGCATCAGATATCC 1743
OY 151 ThrAla-----SerAla 154
|||
|||
|||
Db 1744 ACTTCTCAGCTTCAGATTTCACCCGAATTCGTTCAGACCAAGAGTAGTTAATGA 1803
OY 155 GlyThrProThrProSer-----ProProSerSerTyrLeu 166
|||
|||
|||
Db 1804 GGTGTTCCTGGCAGTCCCTTGCCTTCCTCTCGCCACCTTCTCGCTAC-- 1860
OY 167 ThrIleAspLeuThrAspAspThrGluAspThrHisGlyThrProGlnSerSerThr 186
|||
|||
|||
Db 1861 -----CAGTCAGGTCCCACTCTCTTCACAC--TCGGGCGAGCCAC 1898
OY 187 ProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyMetGluSerProGlnValGlu 206
|||
|||
|||
Db 1899 CCCACACGCGCGCCCTCCAG--GCCCCCTCGCGGACATCCAGACCCCGTCTCACC 1955
OY 207 AlaAspSerGlyAspGlyAspSerSerGluTyr-----Gln 218
|||
|||
|||
Db 1956 CTCTGCTCATGTTCTCCAGCTCTCTCTCTACTATGCTTAACGATGCTTCAGAGG 2015
OY 219 AspGlyLysGluPheGlyIleGlyAspLeuValTyrGlyLysIleLysGlyPheSerTyr 238
|||
|||
|||
Db 2016 GCCTCCAGAGATGTCCCAAGGCGCAGCAGATCTCGAAATCACAGAGTTTCTGTCTGG 2075
OY 239 -----TyrProAlaMetValVal-----Ser 245
|||
|||
|||
Db 2076 GAGGGGTTCATATCCAGTGGCTTGAATTTGATCCCAACCAACCCAGTAGAAGCAGC 2135
OY 246 TrpLysAlaThrSerLysArgGlnAlaMetSerGlyMetArgTyrValGlnTyr----- 263
|||
|||
|||
Db 2136 TACTCTCCAGTAGCAAGACAGACATCTCTCGGGGAA--CGTGTCTATCAGTGTCAAGTG 2194
OY 264 ---PheGlyAsp----- 266
|||
|||
|||
Db 2195 GGGTTCAGAGATTATCCCTTAACATCATAGACCAGGTCTCCAGACAGAACAGATTTG 2254
OY 267 -----GlyLysPheSerSerGluValSerAlaAspLysLeuValAlaLeuGly 281
|||
|||
|||
Db 2255 GAATATCCCGCAGTGGCGCAGTTCTTCTTCTCCCAAGCTGTATTTTCAAATGAG 2314
OY 282 LeuPheSerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLysAla 301
|||
|||
|||
Db 2315 CTGTTGCCATGCTTATTCAGTGCATCTCTACGCTGTAGTCTGCATCGAACAGAG 2374
OY 302 MetTyrHisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSerSerPro 321
|||
|||
|||
Db 2375 CTGTATCCCTTCTATGAGGCTAAAGATT-----CCA 2407
```

```
OY 322 GlyAspSerLeuGluAspGlnLeuLysProMetLeuGluThrAlaHisGlyGlyPheLys 341
|||
|||
|||
Db 2408 GGGTTCAGATCAGAGG--CAGAACTTCTCT-----GCAGGGAAATTA 2448
OY 342 ProThrGlyIleGlyLeuLysProAsnAsnThrGlnProValAlaAsnLysSerLys 361
|||
|||
|||
Db 2449 -----GAAATATTAACCCAAATGAACATCACTTACTTCAAAAGCTGAA 2496
OY 362 ValArgArgAlaGlySerArgLysLeuGluSerArg-----Lys 374
|||
|||
|||
Db 2497 AACAAAGTATATCACACAGTGTCTTGAACATAGAAACAGATGATGATTTAAAGAAA 2556
OY 375 TyrGlnAsnLysThrArgArgArgThrAlaAspSerSerAlaThrSerAspTyrCysPro 394
|||
|||
|||
Db 2557 TTTAAGATGATTTTATGTTACAGCCCAAGTTCTTACTTGAATATATGATCAACTACTA 2616
OY 395 AlaProLysArg-----LeuLysThrAsnCysTyrAsnAsn 406
|||
|||
|||
Db 2617 AACAAAATAGAGAGGAGGAGAAATCAAGAGATTGATCAAAAGCAAAATTGAACCAAGT 2676
OY 407 GlyLysAspArgGlyAspGluAspGlnSerArgGluGlnMetAlaSerAspValAlaAsn 426
|||
|||
|||
Db 2677 GCTAAGGATTTCTTCAATTGAAATAGCAGCAGCAACTGTACAGTGGCAGCAAGCCG 2736
OY 427 AsnLysSerSerLeuGluAspGlyCysLeuSer 437
|||
|||
|||
Db 2737 AATAGCCCGCAGCATTTCCCTTCAATACTTAGT 2769
```

## RESULT 7

```
US-09-105-537-5/C
; Sequence 5, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; US-09-105-537-5
```

## Alignment Scores:

Pred. No.:	0.493	Length:	36778
Score:	137.50	Matches:	111
Percent Similarity:	32.45%	Conservative:	49
Best Local Similarity:	22.52%	Mismatches:	170
Query Match:	3.01%	Indels:	166
DB:	4	Gaps:	21

US-09-720-086-8 (1-853) x US-09-105-537-5 (1-36778)

```
OY 10 GlyGluGluAspAlaGlyGlyArgGluAspSerIleLeuValAsnGlyAlaCysSerAsp 29
|||
|||
|||
Db 31747 GCGCGTGGCGGCGCGGTGTGCGCGGATCGG-----CGGGTGGCGGGCGGTGGG 31697
OY 30 GlnSerSerAspSerProProIleuGluAlaIleArgThrProGluIleArgGlyArg 49
|||
|||
|||
Db 31696 GCAGACCGCGGTGCGGTGAGGTTCACGCGGCGCCAGAGCC-----ATCGGATCAGG 31643
OY 50 ArgSerSerSerArgLeuSerLysArgGluValSerSerLeuLeuSerTyrThrGlnAsp 69
|||
|||
|||
Db 31642 GCCTCGCGGTCCAGG-----TCGTGCATC-----GAC 31616
OY 70 LeuThrGlyAspGlyAspGly-----GluAspGlyAsp 80
|||
|||
|||
```

Db 31615 GCCTCCGGCTCGCACACAGGTCGGCGCGCCCGCTCGAAGCCGCCGGAACCCGGCTCG 31556  
Qy 81 GlySerAspThrProValMetProLysLeuPheArgGluThrArgThrArgSerGluSer 100  
Db 31555 GCCTCG-----ATGCCGGTG-----AGGCCGAGGACGGTGTGAGGACC 31517  
Qy 101 ProAlaValArgThrArgAsnAsnAsnSerValSerSerArgGluArgHisArgProSer 120  
Db 31516 CCCCGCTCCGAGCGCGTCGAGGGCAGTTCGGCAGGCGCGCCGCGCACCCGCCCCCTCC 31457  
Qy 121 ProArgSerThrArgGlyArgGlnGlyArgAsnHisValAspGluSerProValGluPhe 140  
Db 31456 CAGTCCGTCCGGCGCGCTCGCGCGGTGCGAGGTACGCTCCG----- 31415  
Qy 141 ProAlaThrArgSerLeuArgArgAlaThrAlaSerAlaGlyThrProTrpProSer 160  
Db 31414 -----TGGAGGTCCGCGCGCGAGTTCGGCGCGCTCGGTCGTCGAAG 31373  
Qy 161 Pro-----ProSerSerTyrLeuThrIleAspLeuThrAspAspThrGluAspThrHis 178  
Db 31372 ACCGTGTGGCGGGGAGCTGTTCGCGACCCAGCCGCGAGAGCTG-GTTGGGAGCTGCAC 31314  
Qy 179 GlyThrProGlnSerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGly 198  
Db 31313 GGC----- 31311  
Qy 199 GlyMetGluSerProGlnValGluAlaAspSerGlyAspGlyAspSerGluTyr-Gl 218  
Db 31310 -----CGTACGCGAGTCGAAGCCGAGCTCGGTGAAGGACGCGCGGGGCGCACCCG 31260  
Qy 218 nAspGlyLysGluPheGlyIleGlyAspLeuValTrpGlyLysIleLysGlyPheSerTr 238  
Db 31259 GTCCGGGGAGGAA----- 31247  
Qy 238 pTrpProAlaMetValValSerTrpLysAlaThrSerLysArgGlnAlaMetSerGlyMe 258  
Db 31246 -TGCCCGAGTACGGCGCGCGCTGGGTACGGACGAGGTGAGGAGCGCC-----GGCCG 31194  
Qy 258 tArgTrpValGlnTrpPheGlyAspGlyLysPheSerGluVal-----SerAlaAs 275  
Db 31193 GCGC-----TCGGCTCGGGAGCGCGGTGATCGCGGCCAGCGCCGA 31152  
Qy 275 p----- 275  
Db 31151 CGACTGCCCGTCGGCGCACGCGCGCTCGCGGGAGCGGGGCGCACGAGGTGCGGC 31092  
Qy 276 -----LysLeuValAl 279  
Db 31091 GAGCGCTTCGGCGCTTCGGGACGCGCTCGAGCAGAGGCTGGGACGCGACACCGTGA 31032  
Qy 279 aLeuGlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrAr 299  
Db 31031 CGCGGGCGCGAACCGCTCCAGTCGACATCGGCCACCGCGACGAGGCTCTCGCTGGCT 30972  
Qy 299 gLysAlaMetTyrHisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSe 319  
Db 30971 CAGGCGCTTG-----GCCAGTTCTGTCAGGCGCGCTCGGG----- 30935  
Qy 319 rSerProGlyAspSerLeuGluAspGlnLeuLysProMetLeuGluTrpAlaHisGlyGl 339  
Db 30934 -----CTCATCGGACGATCGCGCGACGCTGCCAGTACGCTCGTCCGCG 30891  
Qy 339 yPheLysProThrGlyIleGluGlyLeuLysProAsnAsnThrGlnProValValAsnLy 359  
Db 30890 GCCCGGCGCCATGCGCTCGCGCGCCAGAGGCCCGGCGCGACGAGGTCCCGCTCTCGCC 30831  
Qy 359 sSerLysValArgArgAlaGlySerArgLysLeuGluSerArgLysTyrGluAsnLysTh 379  
Db 30830 CCGGGCGCGCGCGCGCGCGGCGAGCGCGTCG----- 30800  
Qy 379 rArgArgThrAlaAspAspSerAlaThrSerAspTyrCysProAlaProLysArgLe 399  
Db 30799 -----AGGTGGCGGTGGCGCGCGGTAGACCGCCCTGCTCGCGCTGCC----- 30755

Qy 399 uLysThrAsnCysTyrAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGluGl 419  
Db 30754 -CAGACCCCGCGCTTCGAGGAGTA-GAGGACGAAGGCGTCCAGCGGAGTCCCGCGAGCA 30697  
Qy 419 nMetAlaSerAspValAlaAsnAsnLysSerSerLeuGluAspGlyCysLeuSerCys-- 438  
Db 30696 GGTCTCGAGGACCTCG-----CGCCCTCTCTTCCGCCCGCAGATCGGCGGATGCC 30641  
Qy 439 ----GlyArgLysAsnProValSerPheHisProLeu 449  
Db 30640 TCCGGCGCGGTACGTCACGCGGATCGCCCGCGGTG 30604

RESULT 8  
US-09-320-878-19/c  
; Sequence 19, Application US/09320878A  
; Patent No. 6117859  
; GENERAL INFORMATION:  
; APPLICANT: ASHLEY, Gary  
; APPLICANT: BETLACH, Melanie C.  
; APPLICANT: BETLACH, Mary C.  
; APPLICANT: MCDANIEL, Robert  
; APPLICANT: TANG, Li  
; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE  
; FILE REFERENCE: 30622002120  
; CURRENT APPLICATION NUMBER: US/09/320,878A  
; CURRENT FILING DATE: 1999-05-27  
; EARLIER APPLICATION NUMBER: CIP OF 09/141,908  
; EARLIER FILING DATE: 1998-08-28  
; EARLIER APPLICATION NUMBER: CIP OF 09/073,538  
; EARLIER FILING DATE: 1998-05-06  
; EARLIER APPLICATION NUMBER: CIP OF 08/846,247  
; EARLIER FILING DATE: 1997-04-30  
; EARLIER APPLICATION NUMBER: 60/119,139  
; EARLIER FILING DATE: 1999-02-08  
; EARLIER APPLICATION NUMBER: 60/100,880  
; EARLIER FILING DATE: 1998-09-22  
; EARLIER APPLICATION NUMBER: 60/087,080  
; EARLIER FILING DATE: 1998-05-28  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 19  
; LENGTH: 38506  
; TYPE: DNA  
; ORGANISM: Streptomyces venezuelae  
US-09-320-878-19

Alignment Scores:  
Pred. No.: 0.532 Length: 38506  
Score: 137.50 Matches: 111  
Percent Similarity: 32.45% Conservative: 49  
Best Local Similarity: 22.52% Mismatches: 170  
Query Match: 3.01% Indels: 166  
DB: 3 Gaps: 21  
US-09-720-086-8 (1-853) x US-09-320-878-19 (1-38506)  
Qy 10 GlyGluGluAspAlaGlyArgGluAspSerIleLeuValAsnGlyAlaCysSerAsp 29  
Db 29889 GGGCGTGTGGGCGGGTGTGCGCGGATGCG-----CGGGGTGCGCGCGTGG 29839  
Qy 30 GlnSerSerAspSerProIleLeuGluAlaIleArgThrProGluIleArgGlyArg 49  
Db 29838 GCAGACCGCGGTGCGGTGAGGTGTTACGGGGCGCGAGAGCC-----ATCCGGATCAGG 29785  
Qy 50 ArgSerSerSerArgLeuSerLysArgGluValSerSerLeuLeuSerTyrThrGlnAsp 69  
Db 29784 GCCTCGGCTCCAGG-----TCGTCGATC-----GAC 29758  
Qy 70 LeuThrGlyAspGlyAspGly-----GluAspGlyAsp 80  
Db 29757 GCCTCGGCTCCGCACCGAGGTGCGGGCGCGCGCTCCGACCGCGCGAACCCTCGCTCG 29698

QY	81	GLYSERAPRTHPRoValMetProLysLeuPheARGLutHPARGTHARGSerGIUser	100
Db	29697	GGCTCG-----ATGCCGGTG-----AGCGCAGACGGTGTGCAGAAC	296599
QY	101	ProAlaValARGTHARGAsnAsnAsnSerValSerSerARGGLuARGHisARGProSer	120
Db	29658	CCCCCGTCCCGACGCGGTGAGGGGACATTCCGCGCAGGGCCCGCGCACCCGCCCTCC	295939
QY	121	ProARGSerTHARGGLYARGGLNGLYARGAsnHisValARGLuserProValGLuPhe	140
Db	29598	CAGTCCTCGGGGCGCGGTCGCGCGGTGCGAGGTACGCTCG-----	295577
QY	141	ProAlaTHARGSerLeuARGARGAlaTHARGAlaSerIAGLYTHPRoTPProSer	160
Db	29556	-----TGAGAGTCCGCGCGAGAGTCCGCGCGGTGCGGTGTCGAG	295155
QY	161	Pro-----ProSerSerTYRLeuTHILEuPLeuTHIRAsPArTHICuAPRTHHis	178
Db	29514	ACCGTGTGCGCGGGAGCTGTGTGCCAGCACCCGTGAGAGCTG- GTTCCGAGACTGCAC	294566
QY	179	GLYTHPRoGLInSerSerSerTHPRoTYRAlaARGLeuIAGInAsPSeRGLNGInGLY	198
Db	29455	GGC-----	294553
QY	199	GLYMetGLuserPRoGLInValGLuAlaAsPSeRGLYAsPGLYAsPSeRSeRGLUTYR-GL	218
Db	29452	-----CGTCAGGAGTCCGAGCCGAGCTCGTGAAGCAGCCGCGGGGCCACCCG	294020
QY	218	AsPGLYAsGLuPheGLYLeGLYAsPLeuValTPGLYLYsILEuGLYPhSeRTr	238
Db	29401	GTCCGGGAGGAA-----	29389
QY	238	PTTPProAlaMetValAsERTTPLYAlaTHIRSeRLYAsRGLInAlaMetSeRGLYMe	258
Db	29388	-TGCGCGAGTACCGCGCGCGGTGGGTACGAGCGGTGAGAGCGCC-----GGCGG	293366
QY	258	LARGTPVALInTPRheGLYAsPGLYLYsPheSeRGLuVal-----SerAlAs	275
Db	29335	GGCG-----TCGGGCTCGGGAGCGCGGTGATCGCGGCACGCGCGA	292949
QY	275	P-----	275
Db	29293	CGACTGCCGATCGCGGCCACCGCGCGCTCGCGGAGCGGGGACACGCGGTGCGGC	292344
QY	276	-----LYsLeuValAl	279
Db	29233	GAGCGCTGCGCGGCTCCGGGACCGCTCGACAGAGCTGGGACGGGACACCGTGA	291747
QY	279	AlEuGLYLeuPheSeRGLInHisPheAsnLeuAlaTHIRPheAsnLYsLeuValSeRTYR	299
Db	29173	CGCGGCGCGAACCGCTCCACGTGCACATCGGCCACCGCGAGAGGTCTCGTGGCT	291144
QY	299	GLYAlaMetTYRHisAlaEuGLInLYsAlaARGValARGAlaGLYLYsTHIRPheProSe	319
Db	29113	CAGGCGCTTG-----GCCAGTTCGTCCAGGCGCGGTGCGG-----	290777
QY	319	RSerPRoGLYAsPSeRLeuGLuAsPGLInLeuLYsPRoMetLeuGLUTPRAlaHisGLYGL	339
Db	29076	-----CTCATCGACGAGTCCGCGACGCTGCGCAGTACGCGCTCGGC	290333
QY	339	YPheLYsPRoTHIRGLYILEGLInGLYLeuLYsPRoAsnAsnTHIRInProValAlaEnLY	359
Db	29032	GCCCGCGCCATGCGGTGCGCGGCCACAGAGGCCACGAGTGGCGGTGCGC	289737
QY	359	ASerLYsValARGARGAlaGLYSeRARGLYsLeuGLuserARGLYsTYRGLuAsnLYsTh	379
Db	28972	CCGGCGCGCGCGCGCGCGAGCGGTG-----	289424
QY	379	ARGARGARGTHIRAlaAsPAsPSeRAlaTHIRSeRAsPTYRCysPRoAlaPRoLYsARGLe	399
Db	28941	-----AGGTGGCGCTGTGCGCGCGCGTGAAGCGCCCTGTGCGGTGCGC-----	28897
QY	399	uLYsTHIRAsnCYsTYRAsnAsnGLYLYsAsPARGGLYAsPGLuAsPGLInSerARGGLInGL	419

```
Db      28896   -CAGACCCCGGCTTCGAGAGTA-GAGAACAAGGCCTCAAGCGACTGCCGGAGCA    28839
Oy      419     mteAlAserSvPaLaAAnaNySerSeTLeugLiApGlCySLeuSerCy--         438
Db      28838   GGTCTGTGAAGACCCTCG---CGCGCTCTCTTGCCCCCACAAATCGGGCATGTCC    28783
Oy      439     ----GLVArgLysAnProValSerPheHisProIeu          449
Db      28782   TCCGGGCCGTGACCTTCAGCGGATCCGCCGCCGGTGT        28746

RESULT 9
US-08-923-137-1/C
; Sequence 1, Application US/08923137
; Patent No. 6083716
; GENERAL INFORMATION:
; APPLICANT: Wilson, James M.
; APPLICANT: Farina, Steven P.
; APPLICANT: Fisher, Kristina J.
; TITLE OF INVENTION: Chimpazee Adenovirus Vectors
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntrl., P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: United States of America
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,137
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,700
; FILING DATE: 06-SEP-1996
; ATTORNEY/AGENT INFORMATION: *
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GNVPN.02ICPIUSA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 35524 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; US-08-923-137-1

Alignment Scores:
Pred. No.:              .
Score:                  1.25           Length:       35524
Percent Similarity:    133.00         Matches:       149
Best Local Similarity: 28.31%         Conservative:  60
Query Match:           20.90%         Mismatches:   222
                                      Indels:       282
DB:                            3            Gaps:       34

US-09-720-086-8 (1-853) x US-08-923-137-1 (1-35524)
Oy      102 AlAtVaLArgThArGaNaAnASerVerAlsErSeARgLuARGHIsarPROserPro    121
Db      17568 TCAGTTGGCGCTCGAAGCCTGAACCGCCAAGAACGTATTCCCTGGTAGTCCCAGATCGC    17509
Oy      122 ArgeSerTrarrAgLyArvgIngLYArGsEnHiSVaLSpgLUseRPrOVAgiUphEro     141
Db      17508 CGCGGCGTATGATGGGGATCAAGAGGCAAGAAAGCAAACCCCCACAGCCCTTGTGTATAGC    17449
```

Qy	142	AlaThrArgSerLeuArgArgAlaThrAlaSerAlaGlyThrProTrpProSerPro	161
Db	17448	GGCGCGCCGCTGTG-----GGGTGCATCCCGCGCCCAACAT	17410
Qy	162	ProSerSerTyrLeuThrIleAspLeu-----ThrAspAspThr	174
Db	17409	CCCTCTTCTACGGCGGAGTTTCTTCTCGTAGCCAGTAATGGGACACGAAGCGGCA	17350
Qy	175	GluAspThrHisGlyThrProGlnSerSerSerThrProTyr-----	188
Db	17349	AGTGAGGCGCATATCTGCAAGAGGACGCGCAACAGTCGTAAAGTTGTATGCTTGGGT	17290
Qy	189	-----AlaArgLeuAlaGlnAspSerGlnGlnGlyGly-----MetGluSer	202
Db	17289	GGTAGCGTACGGTGGCAACGTCATGGGATCCGCGCGGCCATCGCATACACTCTCGCA	17230
Qy	203	Pro-----GlnValGluAlaAspSer-----GlyAspGlyAspSerSerGluTyrGln---	218
Db	17229	CCAGAGCAGCGGAGTACAGATTGCTGTGGGCGGACGCGACGCGTACACGACTTG	17170
Qy	219	-----AspGlyLysGluPhe-----GlyIleGly-----	226
Db	17169	TAGTGTGTTTGGCGGACGCGGGAAGTACTGCTCCGCTGCGGTAGTAGCGAGTGCCTC	17110
Qy	227	-----AspLeu-----	228
Db	17109	GGTAACCGGAGTTGGAAATAATAGATGGGTGCAGAGCATAAATTTGGCATCAGCAAACTTG	17050
Qy	229	ValTrpGlyLysIleLysGlyPheSerTrpTrp-----	239
Db	17049	CTGGACCGTACTCTCGCGGGATCTTCGAGTGGTGTTCAGCGCAGCTGGAGTTGTAACGG	16990
Qy	240	-----ProAlaMetValValSerTrpLysAlaThrSerLysArgGlnAla	254
Db	16989	GTGCTCACATCCAGGATCGTTTGTACCTCAATGGAGGTGGTTGTGGCTTTGCAGGTT	16930
Qy	255	MetSerGlyMetArgTrpValGln-TrpPheGlyAsp-----GlyLysPheSerGluVa	272
Db	16929	CGGTTTGAACTTCCATGGACCTCGGTTGGAAATCTTGATGTACGGTTTGTACTCCCAAC	16870
Qy	272	1-SerAlaAspLysLeuValAlaLeuGlyLeu-----	282
Db	16869	GTGCGCCACCTGCTTGATGGTCTCACTTTGACCTCAGGTTGTATATCGGATCCACTT	16810
Qy	283	-----PheSerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValS	297
Db	16809	TCATTTTCTCCAGCAGCTCTCCAGCGCTGGCGCTTGGCCACCATCAATTGCTATGTGG	16750
Qy	297	erTyrArgLysAlaMetTyrHisAlaLeu-----	306
Db	16749	GATACAGATCTCCACCTCTCGCTTGATCGCTCTCGTGGAGGACGGGCGCACTTGCT	16690
Qy	307	--GluLysAlaArgValArgAlaGlyLysThrPhePro-----	318
Db	16689	GCAGGTGACTGGCTTGAGCTAGGGTGGGATTTCCCATTTGTCGAAGGNAATGGACACCG	16630
Qy	318	-----	318
Db	16629	CCTCCTCTTGGGACTGGAGCGGAGCGCTTGCCATAAGCAAACTGCCCCAGCGGTCTAG	16570
Qy	319	-----SerSerProGlyAspSerLeuGluAspGlnLeuLysProM	332
Db	16569	CGCGTGTCTCCAGAATATCATCTCCCATACACCTCATATGAGGACCGCTTAAAGACGAG	16510
Qy	332	etLeuGluTrpAlaHisGlyGlyPheLysProThrGlyLleGluGlyLeuLysProAsnA	352
Db	16509	TGCTGGAG--CGCTCGCGCGGGCGTAAGACCACCGTGGTTCAGAGTCTTCGACACAGGCC	16453
Qy	352	snThr-----GlnProValValAsnLysSerLysValArgArgAlaGlySerArgLysLeuG	371
Db	16452	GCACCTTTGGCTGCA-----CGCACTGTCAACACGCGCTTG	16417

Qy	371	luserArgLysTyrGluAsnLysThrArgArgArgThrAlaAspAspSerAlaThrSerA	391
Db	16416	GGCGAACTCGCGCAAAATCCACCAGA	16369
Qy	391	sp-TyrCys-	395
Db	16368	CCTTTTGTGTCCTTTTGTGACCCGCTTGATTTTGGGGGCTTTTTCATCCTTCACCGCGC	16309
Qy	396	ProLysLeuLysThrAsn-	403
Db	16308	GACCGTAGATTTACGCGCGGATGACCTGGAGCAFTCTCTCTGTGTTTGGCCTTGGACA	16249
Qy	404	-----TyrAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGluGln	419
Db	16248	TCCTCGCTGGGACACAAATCCGA-	16225
Qy	420	MetAlaSerAspValAlaAsnAsnLysSerSerLeuGluAspGlyCys-	437
Db	16224	-----GATGCTCATGATCTCTTACAGACGAGGGGACGGGTGCACCGGCACGC	16174
Qy	438	CysGlyArg-----LysAsnProValSerPheHisProLeuPheGluGlyGly	453
Db	16173	GCTGCCGCTGTGGCGTCCGCGACCCAGTACACATTCCTCTCTCGCGTGTGGCCATGT	16114
Qy	454	-----LeuCysGlnThrCysArgAspArgPhe	462
Db	16113	TGGCAATGGCCGCTGCTGCTGCAGTGCCTGCGCGCCTCGGGACCTGCGAGCACTGGCAC	16054
Qy	463	LeuGlu-	464
Db	16053	CTGAAGCGCGCTGTGCTCTAAGCATAGCTCTTCGCCCCACGCGTTTGGCTCTCC	15994
Qy	465	-----LeuPheTyrMetTyrAspAspAspGlyTyrGlnSerTyrCysThr	479
Db	15993	GTAGCAAAAGCTCTTGCTGCTCGCATAGCGGAGTAGCCGCTGGCGCTGGCGATCGTC	15934
Qy	480	ValCysCys-	493
Db	15933	TCCGCTGCTCTTCGGCGAGCATAGCGCGGCGCTCAGCTACTACGCTGTCAATAACTG	15874
Qy	493	rSerCysCysArgCysPheCysValGluCysLeuGluValLeuValGly-	509
Db	15873	CATCCACAGTAGATAGTGTGATAGTAGTAGTAGTAAATACGAGCACTCGGCCACCTCGT	15814
Qy	510	-----ThrGlyThrAlaAlaGluAlaLysLeuGln-	523
Db	15813	CGATCATCATCGACGGTGTCTAGTGGGATACGGCCCTTGGGGCGCCCCATGGAG	15754
Qy	523	rCysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArg-	538
Db	15753	CGCGGAA-----GTGTCCGCGCACTCGCACAGGGTCTGTGTAGAGCGTTTCCGAGCAC	15700
Qy	539	-----ArgLysAspTrpAsnValArgLeuGlnAla-	548
Db	15699	CTCCGTACATCTTGCTTGG-----CGCGCGCAGGCCCCCAACCGGTGTATTACTAGGTG	15646
Qy	549	-----PhePheThrSerAspThrGlyLeuGluTyrGluAlaProly	562
Db	15645	AGATAAGAATGACATTTTTTTTTTTTGTAGAAAGTGGCGCTTGAAGAAGCGCGCGCGC	15586
Qy	562	sLeuTyrProAlaAlaProAlaAlaArgArg	572
Db	15585	ACTATGCCAGGG-----CCTTGTAGAGCTAGG	15558

RESULT 10  
US-08-760-745-4  
; Sequence 4, Application US/08760745  
; Patent NO. 5972658  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Goli, Surya K.  
; APPLICANT: Murry, Lynn E.  
; TITLE OF INVENTION: LUNG GROWTH VARIANT



QY 227 AspleuvlaTrrpGlylsIlelsglyPheSerTrprProAlaMetVal----- 243  
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::  
 Db 355 GACCTGGTGTCGCCCAAGATGAAGGCTACCACACTGGCCGGCCCCGGATTGACGAGATG 414  
 QY 244 -----ValSerTrpLysAlaThrSerLysArgGlnAlaMetSergLysMetArgTrpVal 261  
 :||:|||||:::|||||:::|||||:::|||||:::|||||:::  
 Db 415 CCTGAGCGCTGCCTGAATCATCAACAGCCAACAATACCAA-----GTC 456  
 QY 262 GlnTrpPheGlyAspGlyLysPheSergLysValSerAlaAspLysLeuValAlaLeugly 281  
 ::|||:::|||||:::|||||:::|||||:::|||||:::|||||:::  
 Db 457 TTTTITTTTGGGAGCCCAAGACGGCATTCTGTGGSCCCCAGAAGACCTCTTCCCTTAGCG 516  
 QY 282 LeuPheSergLinhIsPheAsnLeuAlaThrPheAsnLysLeuValSeryTrpArgLysAla 301  
 :||:|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::  
 Db 517 GAATCCAAAGAGAGAAATTGGCAAGCCC-----AACAAAGAGAAAGGTTTCAGCAGGG 570  
 QY 302 MetTrpHisAlaLeuGluLysAlaArgValArgLaGlyLysThrPheProSergSerpPro 321  
 :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::  
 Db 571 CTGTGGAGATCGAAGAACCCCTACTGTCAAGCT---TCCGGCTATCACTCTTCCCG 627  
 QY 322 GlyAspSergLeuGluAspGlnLeuLysPrometLeuGluTrpAlaHisGly--GlyPhe 340  
 :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::  
 Db 628 AAAAAGAGCTGTGTGGAAGAGCCTGAACAGACGCCGAAGCTGCAGAGGGGTGACCGTGAT 687  
 QY 341 LysProThrGlyTrileugLyLeuLysProAsnAsnThrGlnProValValAsnLys--- 359  
 :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::  
 Db 688 AAGAAGGGGAATCGAAGGGCGACGACGACGAGAGAGAGGAGAACCTGGCTATTGATAGCCA 747  
 QY 360 -----SerLysValArgArgAlaGlySerArgLysLeuGlu 371  
 :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::  
 Db 748 GCCAAGAGAGAAAGACGAAAGAGCGTTGAAGAGAGAGACAGGAGACTTGCTGAGAGAC 807  
 QY 372 SerArgLysTyTrgluaenLysThrArgArgArgThrAlaAsp-----AspSeraLatr 389  
 |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::  
 Db 808 TCTCTTAACGTCGCCAAGAGAGCAGAAAACCTGAAGAGAGAGAAAGAGAGCACCCACC 867  
 QY 390 SerAspTyr---CysProAlaProLysArgLeuLysThrAsnCysTyrAsnAsn----- 406  
 :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::  
 Db 868 TTGGAGTTGAGAGGCCCTTCTCTTAGAGGTGAAAAAGAAATGACACCCCTCTGAGCCC 927  
 QY 407 GlyLysAspArgGly-----AspGlnAspGlnSeraArgGluGln 419  
 :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::  
 Db 928 GGCTCTGGCCGGGGGCGCTCCCAAGAGAGAAAGAGAGAGAGATGAAGAGAAAGAGCT 987  
 QY 420 MetAlaSerAsp-ValAlaAsnLysSerSergLeuGluAspGlyCysLeu---Sery 438  
 :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::  
 Db 988 ACCAAGGAAGATGCTGAGGCCCCAGCAGCATCAGAGATGAGAGCCTGTACCAACCAATG 1047  
 QY 438 sGlyArgLysAsnProValSerPheHisPro-----LeuPheGluGl 452  
 :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::  
 Db 1048 TTTCAAGAGAGAGCCCC-----CACCTGTTCTCTGCTGTCTGTGGGTGCTACTGG 1098  
 QY 452 yGlyLeuCysgInThrCybaArg 459  
 :|||:::|||||:::|||||:::|||||:::|||||:::|||||:::  
 Db 1099 GAAACTGGCCATGGCCTGCATAA 1120

RESULT 11  
 US-09-036-987A-1/c  
 ; Sequence 1, Application US/09036987A  
 ; Patent No. 6143526  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Baltz, Richard H.  
 ; APPLICANT: Broughton, Mary C.  
 ; APPLICANT: Crawford, Kathryn P.  
 ; APPLICANT: Madduri, Krishnamurthy  
 ; APPLICANT: Merlo, Donald J.  
 ; APPLICANT: Treadway, Patil J.  
 ; APPLICANT: Turner, Jan R.  
 ; APPLICANT: Walidron, Clive  
 TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
 TITLE OF INVENTION: Production  
 NUMBER OF SEQUENCES: 39  
 CORRESPONDENCE ADDRESS:



ADDRESSEE: Dow Agrosciences LLC Patent Department  
 STREET: 9330 Zionsville Road  
 CITY: Indianapolis  
 STATE: Indiana  
 COUNTRY: USA  
 ZIP: 46268  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/036,987A  
 FILING DATE: 09-MAR-1998  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Stuart, Donald R  
 REGISTRATION NUMBER: 28,479  
 REFERENCE/DOCKET NUMBER: 50,608  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (317)337-4816  
 TELEFAX: (317)337-4847  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 80161 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA (genomic)  
 US-09-036-987A-1

Alignment Scores:  
 Prod. No.: 10.6 Length: 80161  
 Score: 129.50 Matches: 137  
 Percent Similarity: 32.91% Conservative: 72  
 Best Local Similarity: 21.57% Mismatches: 248  
 Query Match: 2.84% Indels: 178  
 DB: 3 Gaps: 30

US-09-720-086-8 (1-853) x US-09-036-987A-1 (1-80161)

QY 20 SerLeuValAsnGlyAlaCysSerAspGlnSerSerAspSerProPheLeuGlu 39  
 DB 29836 ACCGTCCTGATTCACCGCACTACCAACGAAACACCGCCAGACCTCATGCCATTCGCCCG 29777  
 QY 40 AlaileArgThrPro-----Gluile 46  
 DB 29776 GGCATCCGACAAACCGCTCCAGCAACACAGTCGCGCACCTCACCCCATCCAGTGCATC 29717  
 QY 47 ArgGlyArgGSerSerSerArgLeuSerLysArgGluValSerSerLeuSerTyr 66  
 DB 29716 CGACCCCGCCGGAACGACTTCACCGCCCATCCCGCGGAGACCCCGTTGGCGGGAGAA 29657  
 QY 67 ThrGlnAspLeuThrGlyAspGlyAspGlyGluAspGlyAspGlySerAspThrProVal 86  
 DB 29656 CTCAGGAACACCTTCGCGGTGCACATCATCCGTGACGCCACCGCAAGGCGTCCGCA 29597  
 QY 87 MetPro----- 88  
 DB 29596 CTCACCGCCCGCAGTGCTTGCCTCCCGCAGGTGCACACGACCAACGACGACGACGCA 29537  
 QY 89 LysLeuPheArgGluThrArgThr-----ArgSerGluSerProAla----- 102  
 DB 29536 CGTATCCACCGTCCACCGCAGGACCTTCGAACCCGGAACGAATACGCAACCGCCCGGAAAG 29477  
 QY 103 -----ValArgThrArgAsnAsnSerValSerSerArgGluArgHisArgProSer 120  
 DB 29476 AACCGTTCGCGCACCGCGTCCAGCACATGACCTGCACCGCAGGCGCGCTCCGCAA 29417  
 QY 121 ProArg-----SerThrArgGlyArgGlnGly 129  
 DB 29416 CCACGCGGTAGTCTCGGCATTTGGTCCGACGAACACGCCAACCGCGGTGT 29357

QY 130 ArgAsnHisValAspGluSerProValGlu-----PhePro-AlaThrArgSerLeuArgAr 148  
 DB 29356 CGACCGCTGCTCCAGCCAGCCGCTCGAAGACCTCCACGCGACTTCACGACGAAACCG 29297  
 QY 148 g-----ArgAlaThrAlaSerAlaGly----- 155  
 DB 29296 CTGCTCGGATCATCGCCAGCGCTCACCGGGGAAATCCGGAAGAAACCGCGTCCGAA 29237  
 QY 156 -----ThrProTrpProSerProSerSerTyrLeuThrile----- 168  
 DB 29236 CTCACACCGCCCTGAAGAAACCGCCGCGCAGGCATACACGTCGCCGAGGACGATCGGA 29177  
 QY 169 -AspLeuThrAspAspThrGluAspThrHis-----GlyThrProGlnSerSe 184  
 DB 29176 GTCCCGAAGCAACACCGCCATCCAGGTCCCATCCCGCATCCCGGAACTTCGAGATCGC 29117  
 QY 184 rSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyMetGluSer----- 202  
 DB 29116 GTCCACACCGCCGCA-----GCCAGTTGCCACAGTCTCTCCGCGAGGACCTGACC 29063  
 QY 203 -----ProGlnValGluAlaAspSerGlyAspGlyAspSerGluTyrGlnAsp-- 219  
 DB 29062 GGGATAACGACAACCTCATCGCCACGATCGCGATGGGATCGTCTTCTCCGCGGAAACCG 29003  
 QY 220 -----GlyLysGluPheGlyileGlyAsp-LeuValTrpGlyLysileL 234  
 DB 29002 ATCCCTGCGCGCCGCGAGCGTTCGTCTCTTGAAGCATGCTCGCATGCTCGACAAC 28943  
 QY 234 ysGlyPheSerTrpTrpProAlaMetValValSerTyrLysAlaThrSerLysArgGlnA 254  
 DB 28942 TTCTTCGTAACCTGG-----TGGTCAACCGTCATCTCCAGGTCCG 28904  
 QY 254 lametSerGlyMetArgTrpValGlnTrpPheGlyAsp-GlyLysPheSerGluValSer 273  
 DB 28903 CGGTCTCAGACTCGCGTTCGCCCGCGCTGTTTCGACGAGCGCGGTACATCCATCCGA 28844  
 QY 274 AlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAlaThr----- 291  
 DB 28843 TCGATCGAGTCCGCGCCGCAATCG-----TCCGAGGAGATACCGAGTCCGACCGGTGT 28790  
 QY 292 -----PheAsnLysLeuValSerTyrArgLysAla----- 301  
 DB 28789 CCGGCAAGTCGAGCAGCGGACTCACCAACCTGCTCTG-----CGAAGCGCGGAAACCGG 28733  
 QY 302 -----MetTyHisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSer 319  
 DB 28732 ATCGACGTCACTGCGCTCTGATTCGCGCTCGTCTCTCCGGAATGCTCTCTGCG 28673  
 QY 320 SerProGlyAspSerLeuGluAspGlnLeuLysProMetLeuGluTrpAlaHisGlyGly 339  
 DB 28672 GACTCGGGAAGACCTCGCC-----CGCAGTCCGCA 28640  
 QY 340 PheLysProThrGlyileGluGlyLeuLys-----ProAsnAsnThrGlnProVal 356  
 DB 28639 GCCAGTTCGCGAGCGCTGGGATGGCTGAAGATAATTGTCGATGCCAATCGGAGCGGTA 28580  
 QY 357 ValAsnLysSerLysValArgAlaGlySerArgLysLeuGluSerArgLysTyrGlu 376  
 DB 28579 GCTCGATTGAGCGCAACG-----CGAGTTCGACCGCGCTCGAGTTCGAATCCGCTGCC 28523  
 QY 377 AsnLysThrArgArgThrAlaAspSerAlaThrSerAspTyrCysProAlaPro 396  
 DB 28522 TTGAAGGCTCGT-----TCCGATGCAACCGCGCTCGCCGGGTGTCCCAAGCACC 28472  
 QY 397 Lys-----ArgLeuLysThrAsnCysTyrAsnAsnGly---- 407  
 DB 28471 AGACGGGTTTCGCGCGTACCAGCTCCAGGATCAGCTGTTCTCTGTGCGGCTCGGTGCC 28412  
 QY 408 -----LysAspArgGlyAspGluAspGlnSerArgGluGlnMetAlaSerAspVal 424  
 DB 28411 GCGGCAATCTGCGCGCATTCCTCAACGACGCGTCCGCAAGGACTCCCGCTCGGGGTGG 28352  
 QY 425 AlaAsnAsnLysSerSerLeuGluAsp---GlyCysLeuSerCysGlyArgLysAsnPro 443



QY 357 ValAsnLysSerLysValArgArgAlaGlySerArgLysLeuGluSerArgLysTyrGlu 376  
Db 28579 GCTGATTGAGCCGAACG---CGAGTTTCAGCCCGCCCTGCGAGTCCGAATCCCTGTCC 28523  
QY 377 AsnLysThrArgArgArgAlaAspAspSerAlaThrSerArgTyrCysProAlaPro 396  
Db 28522 TTGAAGGCTCGT-----TCCGATGCAACCGCCCTGCGCGGGGTGTCACGACAC 28472  
QY 397 Lys-----ArgLeuLysThrAsnCysTyrAsnAsnGly--- 407  
Db 28471 AGAGCGGTTTCCCGGTACCGATCCAGATCCAGTGTCTTCTGTGCGGCTCGGTGCC 28412  
QY 408 -----LysAspArgGlyAspGluAspGlnSerArgGluGlnMetAlaSerAspVal 424  
Db 28411 CGGACAAATCGTCGCGCATTCACACGACGCGCTCGCAAGGACTCCCGCTCGGGCTGG 28352  
QY 425 AlaAsnAsnLysSerLeuGluAsp---GlyCysLeuSerCysGlyArgLysAsnPro 443  
Db 28351 GCATCGACGAGGAGTTCCTTGACCTCGGCCAGTTCGTTACAGACGCGCCG 28301  
QY 444 ValSerPheHisProLeuPheGluGlyLeuGlyCysGlnThrCysArgAspArgPheLeu 463  
Db 28300 -----GSCCTAGCCGCCACG----- 28286  
QY 464 GluLeuPheTyrMetTyrAspAspAspGlyTyrGlnSerTyrCysThrValCysGly 483  
Db 28285 -----AATCCGCGCTGAAAGCTGCCCAATCGACATCGGCCACT 28247  
QY 484 GlyArgGluLeuLeuLeuCysSerAsnThrSerCysArgCysPheCysValGluCys 503  
Db 28246 ACCAGGAAATTCGTCCTGTTC-----ACGCGCTCGCGGATTCGCGGATCGCGTGC 28193  
QY 504 LeuGluValLeuValGlyThrGlyThrAlaAlaGluAlaLysLeuGlnGluProTyrSer 523  
Db 28192 -----TGCAGGACCATTTGGGGAAGGCGGTATCGCGCGAGCTGGTCT 28151  
QY 524 CysTyrMet-----CysLeuProGlnArgCysHisGly 534  
Db 28150 CCTGTTTTCGCGAGCCATGCGCTCGCGGCCACGCG 28112

## RESULT 13

US-09-491-772-1  
; Sequence 1, Application US/09491772  
; Patent No. 6329172  
; GENERAL INFORMATION:  
; APPLICANT: Pan, Jae Gu  
; APPLICANT: Rhee, Joon Shick  
; APPLICANT: Ahn, Jung Hoon  
; TITLE OF INVENTION: Institute of Science and Technology (KAIST)  
; TITLE OF INVENTION: ABC Transporter Gene Cluster in Pseudomonas fluorescens for  
; FILE REFERENCE: Enhanced Lipase Secretion  
; CURRENT APPLICATION NUMBER: US/09/491,772  
; PRIOR FILING DATE: 2000-01-26  
; PRIOR FILING DATE: 1999-03-13  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: KOPATIN 1.5  
; SEQ ID NO 1  
; LENGTH: 8580  
; TYPE: DNA  
; ORGANISM: Pseudomonas fluorescens  
US-09-491-772-1

## Alignment Scores:

Pred. No.: 0.535 Length: 8580  
Score: 126.00 Matches: 104  
Percent Similarity: 34.84% Conservative: 42  
Best Local Similarity: 24.82% Mismatches: 139  
Query Match: 2.76% Indels: 137  
DB: 4 Gaps: 16

US-09-720-086-8 (1-853) x US-09-491-772-1 (1-8580)  
QY 85 ProValMetProLysLeuPheArgGluThrArgSerGluSerProAlaValArg 104  
Db 5708 CTTATTACACCCAGGCGCTGTTTG---CCAGAGCCAGATCGACATCGCCGCCAAGA 5764  
QY 105 ThrArgAsnAsnAsnSerValSerArg-----GluArgHisArg 118  
Db 5765 AGAAGCCCTCGAAGACAGCAGTTCAGCAGAACCCGACCTGTTCCGAGCAGGCGGCGCA 5824  
QY 119 ProSerPro-----ArgSer 123  
Db 5825 CCGCACCAGATATTCGAAGCCGAGTCGCGTTATGAGCTGCGCCACGCGCCGAGGAGATCG 5884  
QY 124 ThrArg-GlyArgGlnGlyArgAsnHisValAspGluSerProValGluPheProAlaTh 143  
Db 5885 AGCGCTGGACGACGAGGACGCCCTCGCGAGCTGGGGCGTGTGATCGAGTACAGA 5944  
QY 143 rArgSerLeuArgArgAlaThrAlaSerAlaGlyThrProTyrProSerProProSe 163  
Db 5945 GCGTCAACATCGACGACCTGCGCGCTGAGTCC-----TGGCT-TCGCCGCGCTTC 5994  
QY 163 rSer-----TyrLeuThr-IleAspLeuThrAspThrGluAspThrHisG 179  
Db 5995 AGCTTCAGCCGCGCAACTACGACACCTGCGCAGCTGCGCATCAGCAATAACCCACG 6054  
QY 179 ly-ThrProGlnSerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGly 198  
Db 6055 CTGGCATCCAGCGCCAGGCCCTGGAAGTGGCGCTATGAGTGGAGCGCAACCCGCC 6114  
QY 199 Gly-----MetGluSerProGlnValGluAlaAspSerGly 210  
Db 6115 GGGCACCTCGGAAAGTACCGCTATGCGAGTTTCGCGCCAGCAGAGTTCGACACGCGC 6174  
QY 211 AspGlyAspSerGluTyrGlnAspGlyLysGluPheGlyIleGlyAspLeuValTyr 230  
Db 6175 AACACCTACACACCGCTAT---GACACCAACACCATCGCGCTCGAAGTACGCTG--- 6228  
QY 231 GlyLysIleLysGlyPheSerTyrProAlaMetValValSerTyrLysAlaThrSer 250  
Db 6229 -----CCGTTGTATGCGCGTGGCGGCTCTCGCGCTCC 6261  
QY 251 LysArgGlnAlaMetSerGlyMetArgTyrValGlnTyrPheGlyAspGlyLysPheSer 270  
Db 6262 ACTCGCCAGGCCAGCGCGCATGGACGAGCCGCGAGTACGAGCTGGAGGCAAGCGCCG 6321  
QY 271 Glu----- 271  
Db 6322 GAGACCTTGATCGAATGCTGCGCAATTCAGCGCTGTTGTCGGGGTGAGCAAGTTG 6381  
QY 272 -----ValSerAlaAspLysLeuValAla----- 279  
Db 6382 CGCGCTACCAAGAGCGCTGACGTCGCGCGAAGCGTTGGTGTCTACCGCGCAAGC 6441  
QY 280 ---LeuGly-----LeuPheSer 284  
Db 6442 ATCCTTGGCGGTGAGCGGTCATCTTGTGATCGGTCGATCGGACGAGCTGTACAGC 6501  
QY 285 GlnHisPheAsnLeu-----AlaThrPheAsnLysLeuValSerTyrArgLysAlaMet 302  
Db 6502 ACACGCGAGACCTGCGCCAGCGCGGTATGACTACTTGTGCTGGCTGGACCAAGTGCAT 6561  
QY 303 TyrHisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSerSerProGly 322  
Db 6562 TACTAGCCGG-CAATTTGGCGCACCGA-----CCTGGCCCAAGGTAGA 6605  
QY 323 AspSerLeuGluAspGlnLeuLysProMetLeuGluTyrAlaHisGlyGlyPheLysPro 342  
Db 6606 CGAAGCCTTCGGGACCAAGA-GGGCCGAGGCTC----- 6637  
QY 343 ThrGlyIleGluGlyLeuLysProAsnAsnThrGlnProValValAsnLysSerLysVal 362  
Db 6638 -----CCGCTCAGACAAAGCCCTGTTGCAAAACAAACCGCCTG 6676

```

QY 363 ATG-----ArgatAGLYSerhrgLVbLenuIuSerArybLYSTyGlu 376
Db 6677 AGGCAAGCCGACCCCTCGTGCGAGGGGGCTTGCCCGCTTGCGGCAACACCC 6736
QY 377 AsnLYstThrArgArgArgThraIaaspSeralatThrSerApyrCyseProIaPro 396
Db 6737 AATAAAACCAACACAATTCGCGCCAAATCCAC-----AAC 6772
QY 397 LysArgLeuLYstThraSmCyetyrAenAsnGLyLYsAspArgGLyaapGLu 413
Db 6773 AAGAGAGACACAGCAATGGGTGTATTGTACTACAGAACCTCGCACCCGAA 6823

```

[illegible]

```

Db      1777 GCGCTCAGGCGGCCTTCTTCCTCCATGGCTTCATGGACTTTGGACCTTGCGTACTGTGGCGCC 1836
Qy      LysGluGlyAepAspArgProPhePheTrpMetPheGluAenValAlamMetLysVal 704
Db      1837 AAGCGG-----CAGCCCCCTGGGATGGAGCACTACTCGGGC-----TCCTCCAACCTG 1881
Qy      GlyAspLysArgAepIleSerArgPheLeuGluCysAenProvalMetIle----- 721
Db      1882 GGG-----GAGCTGAAGCTCCTGCTAGAG---GAGGCGGTTCATGCTCGACGCCCTC 1929
Qy      722 -----AspAlaIleLysValSerAlaAlaHisArgAlaAatgTyPrPheTrpGly----- 737
Db      1930 AAGGGTGATGTCCTCTCCAGCTCCAGCCAAAGCAGCAAGATGGTGGTGGTGGTGGTGGTGG 1989
Qy      738 -----AenLeuProGlyMetAenArgProVallile-----AlasrLys 750
Db      1990 GCCAGATCAATGTCAGGACCAGAGCGCCCTGGATGCGCCGCCCAAGGAGATGACCACC 2049
Qy      751 AenApLyLysLeuGluLeuGlnAasp-----CysLeuGluTyTrpAenArgIleAla 766
Db      2050 AAGGACAAAATAAGCAGCACGACAAAAGAAGCCCTCATTCCTCTTCTTCAAACAAGACAGCT 2109
Qy      767 LysLeuLysLysValGlnThrIleThrThryLysSerAenSerIleLysGlnGlyLysAen 786
Db      2110 GAAGCTAAATTCCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 2169
Qy      787 GlnLeuPheProValMetAenGlyLysGlu-----AapValLeuTrpCystThrGlu 804
Db      2170 AAG---TTCTGTGTTNCGCACCAATAGGTGGTCTTGGATGCAAT---ACTAAGGAG 2223
Qy      805 LeuGluArgIlePheGlyPheProValHisTyThrAapValSerAenMetGlyArgGly 824
Db      2224 CTTGAGAG-----AAGCGGTGCAGCACATCCGATTCGATGGCTCCACCTCTCTCG 2274
Qy      825 AlaArgGlnLysLeuGlyArgSerTrpSerValPro 837
Db      2275 GCCGACCGGAGACCTCTGCCAGCAGTTTCAGTTGTCCC 2313

RESULT 15
US-08-911-853-29
; Sequence 29, Application US/08911853
; Patent No. 6048710
; GENERAL INFORMATION:
; APPLICANT: Geritsee, Gilsbert
; APPLICANT: Quax, Wilhelmus J.
; TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
; TITLE OF INVENTION: EXPRESSION LEVELS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International
; STREET: 925 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1013
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq For Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/911,853
; FILING DATE:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/699,092
; FILING DATE: 16-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Glaister, Debra J
; REGISTRATION NUMBER: 33,888
; REFERENCE/DOCKET NUMBER: GC361-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-846-7620
; TELEFAX: 650-845-6504

```

```

; INFORMATION FOR SEQ. ID NO.: 29
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 17612 base pairs
;     TYPE: nucleic acid
;     STRADEDNESS: single
;     TOPOLOGY: linear
;
US-08-911-853-29

```

Alignment Scores:	
Pred. No.:	2
Score:	125.50
Percent Similarity:	30.46%
Best Local Similarity:	21.54%
Query Match:	2.75%
BS:	3
Length:	17612
Matches:	140
Conservative:	58
Mismatches:	233
Indels:	222
Gaps:	31

Mon Nov 25 08:18:41 2002

us-09-720-086-8.rni

Page 18

Qy 578 userLeuPheAspGlyIleAlaThrGly 587  
Db 14550 GCGCTTACTGGACGGTGTGGACAGGC 14577

Search completed: November 25, 2002, 03:06:35  
Job time : 473.863 secs

13	126	2.8	8580	4	US-09-491-772-1	Sequence 1, Appl
14	125.5	2.7	2874	4	US-09-179-558-54	Sequence 54, Appl
15	125.5	2.7	17612	3	US-08-911-853-29	Sequence 29, Appl
16	125.5	2.7	17612	4	US-09-479-409-29	Sequence 29, Appl
17	125.5	2.7	17612	4	US-09-479-453-29	Sequence 29, Appl
18	124.5	2.7	4689	4	US-09-105-537-34	Sequence 34, Appl
19	124.5	2.7	42255	4	US-09-199-637A-1	Sequence 1, Appl
20	124	2.7	36778	4	US-09-105-537-5	Sequence 5, Appl
21	124	2.7	38506	3	US-09-320-878-19	Sequence 19, Appl
22	123	2.7	11220	4	US-09-105-537-32	Sequence 32, Appl
23	122	2.7	2151	4	US-09-627-216A-9	Sequence 9, Appl
24	122	2.7	2439	4	US-09-624-693A-18	Sequence 18, Appl
25	122	2.7	5392	2	US-08-403-852D-1	Sequence 1, Appl
26	122	2.7	5392	3	US-08-510-646B-1	Sequence 1, Appl
27	122	2.7	5392	4	US-09-231-818-1	Sequence 1, Appl
28	121.5	2.7	13842	4	US-09-105-537-30	Sequence 30, Appl
29	121.5	2.7	30001	2	US-08-125-468-1	Sequence 1, Appl
30	121.5	2.7	30001	2	US-08-474-933-1	Sequence 1, Appl
31	121.5	2.7	1170	3	US-09-307-621-1	Sequence 1, Appl
32	120.5	2.6	3059	4	US-09-179-558-61	Sequence 61, Appl
33	120	2.6	43280	2	US-08-804-227C-1	Sequence 1, Appl
34	119	2.6	2151	4	US-09-627-216A-7	Sequence 7, Appl
35	118.5	2.6	16836	4	US-09-147-236-1	Sequence 1, Appl
36	118.5	2.6	16836	4	US-09-147-236-10	Sequence 10, Appl
37	118.5	2.6	4403765	4	US-07-101-840A-2	Sequence 2, Appl
38	118.5	2.6	4411529	4	US-09-103-840A-1	Sequence 15, Appl
39	118	2.6	7295	2	US-08-487-826B-15	Sequence 15, Appl
40	117.5	2.6	3688	6	5248670-4	Patent No. 5248670
41	117.5	2.6	34063	4	US-09-453-702B-96	Sequence 96, Appl
42	117.5	2.6	4411529	4	US-09-103-840A-1	Sequence 1, Appl
43	117	2.6	8982	3	US-08-976-255-5	Sequence 5, Appl
44	116.5	2.6	2623	2	US-08-973-675-1	Sequence 1, Appl
45	116.5	2.6	11219	1	US-07-642-734C-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-09-276-531-47  
Sequence 47, Application US/09276531  
Patent No. 6183968  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Hillmezt, Jennifer L.  
APPLICANT: Yue, Henry  
APPLICANT: Reddy, Roopa  
APPLICANT: Guegler, Karl J.  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
NUMBER OF SEQUENCES: 134  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/276,531  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/079,677  
FILING DATE: March 27, 1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

NAME: Lynn E. Murty, Ph.D.  
REGISTRATION NUMBER: 42,918  
REFERENCE/DOCKET NUMBER: PA-0008 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (650) 855-0555  
TELEFAX: (650) 845-4166  
INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2077 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: TESTTUT02  
CLONE: 1271435  
US-09-276-531-47

Alignment Scores:  
Pred. No.: 3,45e-76 Length: 2077  
Score: 903.00 Matches: 183  
Percent Similarity: 89.32% Conservative: 1  
Best Local Similarity: 88.83% Mismatches: 21  
Query Match: 19.78% Indels: 4  
DB: Gaps: 1.

US-09-720-086-8 (1-853) X US-09-276-531-47 (1-2077)  
QY 648 GlySerProCysAspAspLeuSerAsnValAspProAlaArgGlyGlyLeuTyGluGly 667  
DB 3 GCGGCGNANAGCAAGACTNNA-AATGTAATCA-GCCAGAAAGCTTATAGGAC 60  
QY 668 ThrGlyArgLeuPhePheGluPheTyRHisLeuLeuAsnTySerArgProGlyGly 687  
DB 61 AGG--CGGCTCTTCGTAATTTTC-CACCTGCTGAATTTACTCAGCCCAAGAGGCT 116  
QY 688 AspAspArgProPhePheTyRHisLeuLeuAsnValAlaAlaMetCysValGlyAspLys 707  
DB 117 GATGACCGGCGCTTCTTGATGTTGAGATGTTGAGATGTTGAGATGTTGAGATGTTGAG 176  
QY 708 ArgAspIleSerArgPheLeuGluCysAsnProValMetCysAlaIleLysValSer 727  
DB 177 GGGGACATCTCACCGCTTCTGAGAGTATTCAGTATGATGATGATGATGATGATGATGAT 236  
QY 728 AlaAlaHisAValArgTyRHisLeuLeuAsnTySerArgProGlyLysAsnArgIleLys 747  
DB 237 GGTGCTCACAGGCGGCTTCTGAGAGTATTCAGTATGATGATGATGATGATGATGATGAT 296  
QY 748 AlaSerLysAsnAspLeuLeuGluLeuGluLeuGluLeuGluLeuGluLeuGluLeuGlu 767  
DB 297 GATCAAGAAATGATTAATCTCGGCTGAGAGTATTCAGTATGATGATGATGATGATGATGAT 356  
QY 768 LeuLysLysValGlnThrIleThrThrLysSerAsnSerIleLysGlnGlyLysAsnGln 787  
DB 357 TTTAAGAAAGTACACAAATACCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 416  
QY 788 LeuPheProValValMetAsnGlyLysGluAspValLeuTyRHisLeuLeuGluLys 807  
DB 417 CTTTTCCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 476  
QY 808 IlePheGlyPheProValHisTyRHisAspValSerAsnMetCysValGlyAlaArgGln 827  
DB 477 NTCTTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 536  
QY 828 LysLeuLeuGlyArgSerTyRProValIleArgHisLeuPheAlaProLeuLys 847  
DB 537 AAGCTGCTGGAAGGCTCGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 596  
QY 848 AspTyRHisAlaCysGlu 853  
DB 597 GACTACTTGCATGGA 614  
RESULT 2  
US-08-913-832A-1



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 24, 2002, 12:10:37 ; Search time 72.5572 Seconds  
(without alignments)  
3854.741 Million cell updates/sec

Title: US-09-720-086-7  
Perfect score: 4935  
Sequence: 1 MPAMSSSGGDTSSSAARE.....MSVPVIRHLFAPLKEYFACV 912

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
MODELS=frame+ p2n.model -DEV=xlh  
-O=/cg2n\_1/USPTO\_epool/US09720086/runat\_18112002\_092957\_29457/app\_query.fasta\_1.4252  
-DB=Issued\_Patents\_NA -QPM=fastap -SUFFIX=eni -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTPM=ptc -NORM=ext -HEAPSIZE=500 -MTLEN=0 -MAXLEN=2000000000  
-USER=US09720086 @CGN\_1\_1\_108 @runat\_18112002\_092957\_29457 -NCP=6 -ICP=3  
-NO\_XUPXY -NO\_MAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_NA: \*  
1: /cg2n\_6/ptodata/1/ina/5A\_COMB.seq: \*  
2: /cg2n\_6/ptodata/1/ina/5B\_COMB.seq: \*  
3: /cg2n\_6/ptodata/1/ina/5A\_COMB.seq: \*  
4: /cg2n\_6/ptodata/1/ina/5B\_COMB.seq: \*  
5: /cg2n\_6/ptodata/1/ina/PCTUS\_COMB.seq: \*  
6: /cg2n\_6/ptodata/1/ina/Backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	748	15.2	2077	4	US-09-276-531-47	Sequence 47, Appl
2	182.5	3.7	15872	4	US-09-105-537-1	Sequence 1, Appl
3	182.5	3.7	43280	2	US-08-804-227C-1	Sequence 1, Appl
4	172.5	3.5	12001	1	US-08-458-568A-11	Sequence 11, Appl
5	171	3.5	4852	1	US-07-853-913-3	Sequence 3, Appl
6	166	3.4	3014	1	US-08-629-939-1	Sequence 1, Appl
7	166	3.4	3014	1	US-08-759-873-1	Sequence 1, Appl
8	166	3.4	3251	4	US-09-085-199B-6	Sequence 6, Appl
9	166	3.4	35060	3	US-08-814-095-7	Sequence 7, Appl
10	163.5	3.3	4350	4	US-09-295-593-37	Sequence 37, Appl
11	161	3.3	7791	3	US-08-949-386-23	Sequence 23, Appl
12	161	3.3	7791	3	US-08-450-562-23	Sequence 23, Appl

13	161	3.3	7791	4	US-08-984-709A-23	Sequence 23, Appl
14	161	3.3	7791	4	US-08-450-272-23	Sequence 23, Appl
15	161	3.3	7808	3	US-08-949-386-22	Sequence 22, Appl
16	161	3.3	7808	3	US-08-450-562-22	Sequence 22, Appl
17	161	3.3	7808	4	US-08-984-709A-22	Sequence 22, Appl
18	161	3.3	7808	4	US-08-450-272-22	Sequence 22, Appl
19	159	3.2	2277	1	US-08-676-967-5	Sequence 5, Appl
20	159	3.2	2277	1	US-08-676-974-5	Sequence 5, Appl
21	159	3.2	2277	2	US-09-098-487-5	Sequence 5, Appl
22	159	3.2	8438	1	US-07-945-283-1	Sequence 1, Appl
23	158.5	3.2	2680	4	US-09-063-035-1	Sequence 1, Appl
24	158.5	3.2	9551	1	US-08-056-200-93	Sequence 93, Appl
25	158.5	3.2	9551	2	US-08-800-644-93	Sequence 93, Appl
26	157	3.2	20235	3	US-07-642-734C-3	Sequence 3, Appl
27	157	3.2	20235	3	US-08-439-009A-3	Sequence 3, Appl
28	156.5	3.2	1736	4	US-09-162-524-2	Sequence 2, Appl
29	156.5	3.2	7808	2	US-08-149-097D-22	Sequence 2, Appl
30	155	3.1	1681	4	US-09-434-288-7	Sequence 7, Appl
31	155	3.1	5468	4	US-09-535-008-66	Sequence 66, Appl
32	155	3.1	5471	4	US-09-535-008-1	Sequence 1, Appl
33	155	3.1	5471	4	US-09-535-008-74	Sequence 74, Appl
34	155	3.1	5480	4	US-09-535-008-70	Sequence 70, Appl
35	155	3.1	5564	4	US-09-535-008-68	Sequence 68, Appl
36	155	3.1	5576	4	US-09-535-008-64	Sequence 64, Appl
37	155	3.1	5573	4	US-09-535-008-76	Sequence 76, Appl
38	155	3.1	5576	4	US-09-535-008-72	Sequence 72, Appl
39	155	3.1	13842	4	US-09-105-537-30	Sequence 30, Appl
40	155	3.1	36778	4	US-09-105-537-5	Sequence 5, Appl
41	155	3.1	38506	3	US-09-320-878-19	Sequence 19, Appl
42	154	3.1	5467	1	US-07-745-206A-12	Sequence 12, Appl
43	154	3.1	5467	2	US-08-311-363-12	Sequence 12, Appl
44	154	3.1	11907	4	US-08-061-376-4	Sequence 4, Appl
45	154	3.1	14255	1	US-08-320-559-1	Sequence 1, Appl

## ALIGNMENTS

RESULT 1  
US-09-276-531-47  
Sequence 47, Application US/09276531  
Patent No. 6183968  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Yue, Henry  
APPLICANT: Reddy, Roopa  
APPLICANT: Guegler, Karl J.  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION  
NUMBER OF SEQUENCES: 134  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/276,531  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/079,677  
FILING DATE: March 27, 1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

NAME: Lynn E. Murry, Ph.D.  
 REGISTRATION NUMBER: 42,918  
 REFERENCE/DOCKET NUMBER: PA-0008 US  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (650) 855-0555  
 TELEFAX: (650) 845-4166  
 INFORMATION FOR SEQ ID NO: 47:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2077 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 IMMEDIATE SOURCE:  
 LIBRARY: TESTTUT02  
 CLONE: 1271435  
 US-09-276-531-47

Alignment Scores:  
 Pred. No.: 2,1e-46 Length: 2077  
 Score: 748.00 Matches: 151  
 Percent Similarity: 81.95% Conservative: 17  
 Best Local Similarity: 73.66% Mismatches: 36  
 Query Match: 15.16% Indels: 4  
 DB: 4 Gaps: 1

US-09-720-086-7 (1-912) x US-09-276-531-47 (1-2077)

QY 707 GlySerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeuTyrGluGly 726  
 Db 3 GCGGCGNANTGCAACGATCTNNAAT-GTGAATCA-GCCAGGAAGGCTGTATGAGGAC 60  
 QY 727 ThrGlyArgLeuPheGluPheTyrArgLeuLeuHisAspAlaArgProLysGluGly 746  
 Db 61 AGG---CGGCTCTTCTTGAATTTTC-CACCTGCTGATTTACTACGCCCCCAAGAGGGT 116  
 QY 747 AspAspArgProPhePheTrpLeuPheGluAsnValValAlaMetGlyValSerAspLys 766  
 Db 117 GATGACCGGNCCTTCTTCTGGATGTTTGAAGATGTTGTAGNCTGCTCGCGATTCCGGCAGA 176  
 QY 767 ArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLysGluValSer 786  
 Db 177 GGGGACATCTACGGTCTCTGGAGTGTATCCAGTGATGATGATGATGATGATGATGATGAT 236  
 QY 787 AlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMetAsnArgProLeuAla 806  
 Db 237 GCTGCTCAGGGCCCGGATCTTCTGGGCAACTCTCCCGGATGAACAGGCCCGGTGATA 296  
 QY 807 SerThrValAsnAspLysLeuGluLeuGluGlnCysLeuGluHisGlyArgIleAlaLys 826  
 Db 297 GCATCAAGAATGATAAACTCGNGCTGCAGGACTGCTTGGAAATACAAATAGGATAGCCAAAG 356  
 QY 827 PheSerLysValArgThrIleThrArgSerAsnSerIleLysGlnGlyLysAspGln 846  
 Db 357 TTAAGAAAGTACAGACATAACCAACAGTCGAACCTCGATCAACAGGGGAAACCA 416  
 QY 847 HisPheProValPheMetAsnGluLysGluAspIleLeuTrpCysThrGluMetGluArg 866  
 Db 417 CTTTTCCTGTTGTCATGATGCAAGAAAGATGTTTNGTGGTGCACCTGAGCTCGAAAGG 476  
 QY 867 ValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeuAlaArgGln 886  
 Db 477 NCTTTGGCTTTCTCTGTGACTACACAGAGTGTCCCAACATGGCCCGTGTGCCGCCAG 536  
 QY 887 ArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAlaProLeuLys 906  
 Db 537 AAGCTGCTGGGAGGTCTTGAGCGTGCTGTATCCGACACCTCTTCGCCCCCTCTGAAG 596  
 QY 907 GluTyrPheAlaCys 911  
 Db 597 GACTACTTTTGCATGT 611  
 RESULT 2  
 US-09-105-537-1/c

; Sequence 1, Application US/09105537A  
 ; Patent No. 6265202  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sherman, D.H.  
 ; APPLICANT: Liu, H.  
 ; APPLICANT: Xue, Y.  
 ; APPLICANT: Zhao, L.  
 ; TITLE OF INVENTION: DNA encoding methymycin and pikromycin  
 ; FILE REFERENCE: 600,438US1  
 ; CURRENT APPLICATION NUMBER: US/09/105,537A  
 ; CURRENT FILING DATE: 1998-06-26  
 ; NUMBER OF SEQ ID NOS: 43  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 1  
 ; LENGTH: 15872  
 ; TYPE: DNA  
 ; ORGANISM: Streptomyces venezuelae  
 ; US-09-105-537-1

Alignment Scores:  
 Pred. No.: 0.00274 Length: 15872  
 Score: 182.50 Matches: 137  
 Percent Similarity: 32.84% Conservative: 64  
 Best Local Similarity: 22.39% Mismatches: 209  
 Query Match: 3.70% Indels: 202  
 DB: 4 Gaps: 33

US-09-720-086-7 (1-912) x US-09-105-537-1 (1-15872)

QY 38 ArgGlnGluProSerThrThrAlaArgLysValGlyArgProGlyArgLysArgLysHis 57  
 Db 10619 CGGAGATCAACCCAGCCGCGGCGCTTCA-----CCGCGCGCAGCAGACGCGGT 10569  
 QY 58 Pro-----ProValGluSerGlyAspThrProLysAspProAlaValIleSerLys 74  
 Db 10568 CCTCGGCGCGCCAGCAGTGCAGACCTGCTCTCGCGGCCCGCA-----CGG 10518  
 QY 75 SerProSerMetAlaGlnAspSerGlyAlaSerGluLeuLeu-----Prohndly 91  
 Db 10517 AGACCATCCGCGCGCGCGCGCGCTGCATCAGCCTCCCGCGCGCGCCACCGCG 10458  
 QY 92 AspLeuGluLysArgSerGluProGlnProGluGluGlySerProAlaGlyGlyGlnLys 111  
 Db 10457 CCGAGCGCTCGTGCAGATCGAGCACACCCCGCGCGCGCGCA---GTTCCGCCA 10401  
 QY 112 GlyGlyAlaProAlaGluGlyGluGlyAlaAlaGluThrLeuProGluAlaSer----- 129  
 Db 10400 CCGAGTGCCTCCAGCAGCAGTCCGCGCGCATCCCCAGTGTCTCCAGCAGCGCGTGAGGG 10341  
 QY 130 -----ArgAlaValGluAsnGlyCysCysThrProLysGluGlyArgGlyAla 145  
 Db 10340 CCACCTCGACGGCGAAGAGCGCGCTGGGTGT-----ACTCGTCCGCTCCAGGAGTG 10287  
 QY 146 ProAlaGluAla-----GlyLys-GluGlnLys-----GluThrAs 157  
 Db 10286 CCGCTTCGGCGAGCGCGGTCTCGGCGCTCCAGCAGCGCGCGCGCGCGCGCGCGAGCGG 10227  
 QY 157 ntleGluSerMetLys-----MetGluGlySerArgGlyArgLeuArgGlyGlyLeu-- 174  
 Db 10226 GGTTCGAGACGCTCCCTCACCCTCGTCCAGGCGCGCGCGCGCGCGCGCGCGCGCGATGCA 10167  
 QY 175 --GlyTrpGluSerSerLeuArgGlnArgPro-----184  
 Db 10166 GGTCTGGGCCCATCCCGCGCGCTGGTGGTCCCTCCCGGTGAAGAGGAGGCATCTCTGC 10107  
 QY 185 -----MetProArgLeuThrPheGlnAlaGlyAspProTyrTyrIleSerL 200  
 Db 10106 CGTCCCGGTGCGGACTCTCCGCGCACACCGCGCGCGCGCGCGCGCGCGCGCGCGGT 10047  
 QY 200 yserg-----LysArgAspGluTrpLeuAlaArgTrpLysArg 213  
 Db 10046 CCAGGCGCGCGCGAGTTCGCGCGCGCTCCGAGCGCGATCAGCACGCGCGCGGTGG----- 9994

OY	213	1aa1agluhlyslalysalalysal11lealaclymetasna1alvalgluhlunsglnclyp	233
Db	9993	-----TCMAAGGCTGCGCGCGCGCGCGAGAGAGAGAGCGATGTCA CGCGGCGGCGCC	9939
OY	233	rogllygluserglnlyslvalglunl-----	Alas 243
Db	9938	CGGAGGTGCGGAGAGGTGTCGCGAGCCGCGGCGCTGGAGCCGAGGCGCGCTCGT	9879
OY	243	erProProAlaValGlnGlnProThrAspProAlaSerProThraValAlaThrPro-	262
Db	9878	CGCGTCCGAGACCAACCAGGCCAGGAGCCCTTCGCCCTTCCTTCGCGCGCGGAGCGCG	9819
OY	263	-----GluProValGlySerAspAlaGlyAspLysAsnAlaThr-----	275
Db	9818	CGGAGGCTCCCGCGCGGCTTCGCGCGGAGCGGAGCGGTGCTTCGAGAGACAT	9759
OY	276	-----LysAla-----	277
Db	758	GAGCATGTGTGCGCGCCAGCGCCGAGAGCGGACACCCCGGTCCGCGGCGCGCGCTCC	9699
OY	278	-----GlyAspAspGlu-----ProGluTyrGlnAspGlyArg	289
Db	9698	GCTGCCACGCGACCTCGAGAGTGAGACGCCGTACCCGCGCGGAGCTCCAGTCG-----G	9645
OY	289	lyPheGlyTylegIygluleuValTropGlyLysLeuArgGlyPheSerTrpProGlyA	309
Db	9644	CATGCGGGGTGCGCTGTCGCGGTG--AGGATGCGGGGACAGCAG-----CCGTGCC	9594
OY	309	rgIleValSerTrpTrp-----MetThrGlyArgSerArgAlaAlaGluGlyThra	326
Db	9593	GC-----AGCGCTGAGCATTTGATGACGCCGCGCACCGCGCGCGCTGAGCAT	9548
OY	326	rgTrpValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLysLeuMetP	346
Db	9539	GCCCGATGTGACTTCAGCGAGCCGAGCCACGACGAGGTGTGCGCCGGAACGTTCCGTG	9488
OY	346	roleuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnProMetTyrA	366
Db	9479	CGAGCGCGCGAGAGCGCGCCCGCGCTCGATGAGGTGCGCGAGCGCGGTGCGGTG---	9424
OY	366	rgLysAlaIeTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeuPheProV	386
Db	9423	-----CCGTGCGCTTCGACCGCGTCAACGCTTCGCGG-----	9391
OY	386	alCysHisAspSerAspLysSerAspThrAlaLysAlaVal-GluValGlnAsnLysPro	405
Db	9344	CTGGCGCTGTGGGCGCGGTCAAGCCCTGTGAGAGCAACGTCTGCTGTGACCGCGGTGCGG	9285
OY	413	-----GlyPheGlnProSerGlyProLysGlyLeuGlnPro-----	424
Db	9284	TGACCAAGCGCGAGAGACGTGTGCGCGCGCGCGCGCTTCACAGCCGCTCCAGAGCA	9225
OY	424	-----	424
Db	9224	GCACGCGAGCGCTCGCGCCAGCCGCGTGCCTGCGCGCGCGCGCGAGAACGCTTGCAGC	9165
OY	425	-----Pro-----	425
Db	9164	GAGCGTCCGCGCGAGCCCTTCTCGCGGAGAACTCCACGAACGCGCGGATCGGCA	9105
OY	426	GluGluGlnLysAsnProTyrLysGluValTyrThrAspMetTrpValGluProGluAla	445
Db	9104	TCACGCGCACCCCGCTCCGAGGCGAGCCGCACT-----CGCGGTGC	9066
OY	446	AlaAlaTyrAlaProPro-----ProProAlaLysLysProArgLysSerThrAla	462
Db	9059	GCAGCGCGCGTACCGCCAGATGAGCGCCACACGAGAGAGAGACGAGCGCTCT-----	9005
OY	463	GluLysProLysValLysGluLeileAspGlnAsyThrArgGluArgLeu--ValTyr	481

Db 9005 -----CCACCCGTCATGAGCCGTCCTTCGATCCGACCCGAGCCGCTAGAGATGCGGCCGACA 8952  
Oy 482 GluValArgGlnLysCysArgAsnIleGlu 491  
Db 8951 TCACCGCGTCGGCGCTGCCGCTACGAGAT 8922

RESULT 3  
US-08-804-227C-1/C  
; Sequence 1, Application US/08804227C  
; Patent No. 5876991  
; GENERAL INFORMATION:  
; APPLICANT: Dehoff, Bradley S.  
; APPLICANT: Kuhstoss, Stuart A.  
; APPLICANT: Rostock, Paul R., Jr.  
; APPLICANT: Sutton, Kimberly L.  
; TITLE OF INVENTION: POLYMETIDE SYNTHASE GENES  
; NUMBER OF SEQUENCES: 15  
; ORDER INFORMATION:

ADDRESS: THOMAS G. PLANT 1501  
STREET: LILLY CORPORATE CENTER  
CITY: INDIANAPOLIS  
STATE: IN  
COUNTRY: USA

```

; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: Floppy disk
;
; COMPUTER: IBM Compatible
;
; OPERATING SYSTEM: MS-DOS
;
; SOFTWARE: ASCII(DOS) Text only
;
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/08/804,227C  
FILING DATE: February 21, 1997

```

; CLASSIFICATION: 435
;
; ATTORNEY/AGENT INFORMATION
;

```

NAME: Plant, Thomas, G.  
REGISTRATION NUMBER: 35,784  
REFERENCE/DOCKET NUMBER: X-82311

TELECOMMUNICATION INFORMATION  
TELEPHONE: 317-276-2459  
INFORMATION FOR SEQ ID NO: 1:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 43280 base pairs
; TYPE: nucleic acid
;

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;

```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: 816..14234
;

```

FEATURE:	CDS
NAME/KEY:	
LOCATION:	14351..199

FEATURE:	CDS
NAME/KEY:	
LOCATION:	20010..311

FEATURE:	CDS
NAME/KEY:	31232..360
LOCATION:	

```

FEATURE:
NAME/KEY: CDS
LOCATION: 36249, 41774

```

US-08-804-227C-1  
Alignment Scores

Pred. No.:  
Score:  
Percent Similarity:

Best Local Simil  
Query Match:  
DB:

US-09-720-086-7 (1-912) x US-08-804-227C-1 (1-43280)

```

Db      9005 -----CCACCCTCATGACGCGTCCCTTCACAGACGACGCGCGTAGAGAT
Oy      482 GluValAlaArgIlnIysCysArgAsnIleGlu 491
Db      8951 TCACCGCGCTGGCGGCTGCGGCTCAGCAGAT 8922

RESULT 3
US-08-804-227C-1/c
; Sequence 1, Application US/08804227C
; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sulton, Kimberley L.
; TITLE OF INVENTION: POLYMERIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ. ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43280 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 816..14234
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14351..19945
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20010..31199
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31232..36067
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36249..41774
; US-08-804-227C-1

Alignment Scores:
Pred. No.: 0.011 Length: 43280
Score: 182.50 Matches: 137
Percent Similarity: 32.84% Conservative: 64
Best Local Similarity: 22.39% Mismatches: 209
Query Match: 3.70% Indels: 202
Db: 2 Gaps: 33

US-09-720-086-7 (1-912) x US-08-804-227C-1 (1-43280)

```

QY 38 ArgGlnGluProSerThrThrAlaArgLysValGlyArgProGlyArgLysHis 57  
DB 38481 CGGAGTACACACCGCGGGCGCTTCA-----CGCGCGCAGCAGCGCGT 38431  
QY 58 Pro-----ProValGluSerGlyAspThrProLysAspProAlaValLysHis 74  
DB 38430 CCG 38380  
QY 75 SerProSerMetAlaGlnAspSerGlyAlaSerGluLeuLeu-----ProAsnGly 91  
DB 38379 AGACCATCG 38320  
QY 92 AspLeuGluLysArgSerGluProGlnProGluGluGluSerProAlaGlyGlyGlnLys 111  
DB 38319 CGCAGCGGTGCTCGAGATCGAGCACACCGCGCGCGCGCGCGCGCGCGCG 38263  
QY 112 GlyGlyAlaProAlaGluGlyGluGlyAlaAlaGluThrLeuProGluAlaSer----- 129  
DB 38262 CCGAGTGCAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38203  
QY 130 -----ArgAlaValGluAsnGlyCysCysThrProLysGluGlyArgGlyAla 145  
DB 38202 CCACCTCGACGCGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 38149  
QY 146 ProAlaGluAla-----GlyLys-GluGlnLys-----GluThrAs 157  
DB 38148 CCGCTTCG 38089  
QY 157 nileGluSerMetLys-----MetGluGlySerArgGlyArgLeuArgGlyLeu-- 174  
DB 38088 GGTGACAGCGGTTCGTACCTCGTCAGGCGCGCGCGCGCGCGCGCGCGCGCG 38029  
QY 175 --GlyTrpGluSerSerLeuArgGlnArgPro----- 184  
DB 38028 GGTGTCGTGCGCATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 37969  
QY 185 -----MetProArgLeuThrPheGlnAlaGlyAspProTrpTrpTrpSerL 200  
DB 37968 CGTCCCGGTGCGGACTCCCGCGCACCGCGCGCGCGCGCGCGCGCGCGCG 37909  
QY 200 ysArg-----LysArgAspGluTrpLeuAlaArgTrpLysArgG 213  
DB 37908 CCAGCG 37856  
QY 213 luAlaGluLysLysAlaLysValIleAlaGlyMetAsnAlaValGluGluAsnGlnGlyP 233  
DB 37855 -----TCAAAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 37801  
QY 233 roGlyGluSerGlnLysValGlu-----AlaAs 243  
DB 37800 CGGGGTGCGGAGAGGTGTCGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 37741  
QY 243 erProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrPro- 262  
DB 37740 CGCGTCCGAGAGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 37681  
QY 263 -----GluProValGlySerAspAlaGlyAspLysAsnAlaThr----- 275  
DB 37680 CGGGGCGCTCCCGCGCGGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 37621  
QY 276 -----LysAla----- 277  
DB 37620 GGGCATGTCG 37561  
QY 278 -----GlyAspAspGlu-----ProGluTrpGluAspGlyArgG 289  
DB 37560 GCTGCCACCGCACCTCGGAGGTGAGCAGCGCGTACCAGCGCGAGCTCCAGTCG-----G 37507  
QY 289 lypheGlyIleGlyGluLeuValTrpGlyLysLeuArgGlyPheSerTrpTrpProGlyA 309  
DB 37506 CATCGGGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 37456

QY 309 rgIleValSerTrpTrp-----MetThrGlyArgSerArgAlaAlaGluGlyThrA 326  
DB 37455 GC-----AGCGCTGGAGCATCTTGATGAGCGCGCGCGCGCGCGCGCAT 37402  
QY 326 rgTrpValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLysLeuMetP 346  
DB 37401 GCCGATGTTGGACTTCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 37342  
QY 346 roLeuSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnProMetTyrA 366  
DB 37341 CGGAGCG 37286  
QY 366 rgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeuPhePro 386  
DB 37285 -----CCGTGCGCTCGACCGCGCTCCACTCTCCCGG----- 37253  
QY 386 alCysHisAspSerAspGluSerAspThrAlaLysAlaVal-GluValGlnAsnLysPro 405  
DB 37252 -----GACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 37207  
QY 406 MetIleGluTrpAlaLeuGly----- 412  
DB 37206 CTGGCGCGGTGGCGCGGTGAGCGCGGTGGAGGACCGCTCTGTTGACCGCGGTGCGCG 37147  
QY 413 -----GlyPheGlnProSerGlyProLysGlyLeuGluPro----- 424  
DB 37146 TCACCGAGCGCGAGGACCGGTGTCGCCCGCGCGCGCGCGCGCGCGCGCG 37087  
QY 424 ----- 424  
DB 37086 GCACCGCGAGCGCTCGGCCCGCGCGGTGCGGTGCGCGCGCGCGCGCGCGCG 37027  
QY 425 -----Pro 425  
DB 37026 GCGCGTGGCG 36967  
QY 426 GluGluGluLysAsnProTrpLysGluValTyrThrAspMetTrpValGluProGluAla 445  
DB 36966 TCACCG 36922  
QY 446 AlaAlaTyrAlaPro-----ProProAlaLysLysProArgLysSerThrAla 462  
DB 36921 GCAGCGCGCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 36868  
QY 463 GluLysProLysValLysGluIleAspGluArgThrArgGluArgLeu---ValTyr 481  
DB 36867 -----CCACCGCTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 36814  
QY 482 GluValArgGlnLysCysArgAsnIleGlu 491  
DB 36813 TCACCGCGTGGCGCGTGGCGCGTGGCGCGTGGCGCGTGGCGCGTGGCG 36784  
RESULT 4  
US-08-458-568A-11  
; Sequence 11, Application US/08458568A  
; Patent No. 5821339  
; GENERAL INFORMATION:  
; APPLICANT: Schaffer, Priscilla A.  
; APPLICANT: Yeh, Lily  
; TITLE OF INVENTION: Compositions and Methods for Treatment of Herpesvirus  
; TITLE OF INVENTION: Infections  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339ris  
; STREET: One Liberty Place, 46th floor  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS



Db 7100 CGGGGGGACCCCGCCCAAGGGGGGGGGGGGGGGGGGGTAAAGAAAGTGAAGAACGGCGAAGCG 7159  
Qy 477 -----GluArgLeuValTyrGluValArgGlnLysCysArgAsn 489  
Db 7160 TTGCGACTTCGTCCCAATATATATATATTATTAGGGCGAAGTGGCGAC 7209  
RESULT 5  
US-07-853-913-3  
; Sequence 3, Application US/07853913  
; Patent No. 5338839  
; GENERAL INFORMATION:  
; APPLICANT: McKay, Ronald D.G.  
; TITLE OF INVENTION: Nestin Expression As An Indicator of  
; TITLE OF INVENTION: Neuroepithelial Tumors  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Millicia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/853,913  
; FILING DATE: 19920319  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/660,412  
; FILING DATE: 22-FEB-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/603,803  
; FILING DATE: 25-OCT-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/201,762  
; FILING DATE: 02-JUN-1988  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/180,548  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia  
; REGISTRATION NUMBER: 32,227  
; REFERENCE/DOCKET NUMBER: MIT-4641AAAA  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-861-6240  
; TELEFAX: 617-861-9540  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4852 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-07-853-913-3  
Alignment Scores:  
Pred. No.: 0.00378 Length: 4852  
Score: 171.00 Matches: 126  
Percent Similarity: 32.32% Conservative: 75  
Best Local Similarity: 20.26% Mismatches: 191  
Query Match: 3.47% Indels: 231  
DB: 1 Gaps: 27  
US-09-720-086-7 (1-912) x US-07-853-913-3 (1-4852)  
Qy 6 SerSerGlyProGlyAspThrSerSerSerAlaAlaGluArgGluGluAspArgLysAsp 25  
Db 2689 AGCATGGAACTGGAGATTTCAGATCTCCAGAGGAGTAGACAAAGGAAGTCAAAGGAAT 2748

Qy 26 GlyGluGluGlnGluGluProArgGlyLysGluGluArgGlnGluProSerThrThrAla 45  
Db 2749 CTGGAGAG---GAAGAGAACCTGGGAAAGGGAGAGACCAAGAG---TCACGTG 2796  
Qy 46 ArgLysValGlyArgProGlyArgLysArgLysHisProProValGluSerGlyAspThr 65  
Db 2797 AGGTCTCTGGAGGAGGAGGAGGAGGAGCTGCCG-----CACTCTGCAGATGTG 2844  
Qy 66 ProLysAspProAlaValLleSerLysSerProSerMetAlaGlnAspSer----- 82  
Db 2845 CAGAGGTGGGAAGATACGGTGGGAAGAGGACCAAGAACTGGCTCAGGAAAGCCCTCTCGG 2904  
Qy 82 ----- 82  
Db 2905 ATGGCTGGAGTGAAATAAGGATAGGCGAGAGCTGAATCTAAGGAGCAGGATGCTTC 2964  
Qy 83 ---GlyAlaSerGluLeuLeuProAsnGlyAspLeuGluLysArgSerGlu----- 98  
Db 2965 ACTGGGAAGGAGGTGGTAGAGAGGAGGCTGAATGCCACAGAGGAGGTCTGGTTC 3024  
Qy 99 -----ProGlnProGluGlu----- 103  
Db 3025 CCAGCGAGGGGCAACCCAGAGAAACCTTGAGCCCAAGAGCAGAGAGGCTTGGTTGAGGGA 3084  
Qy 104 GlySerProAlaGlyGlnLysGlyGlyAlaProAlaGluGlyGluGly----- 120  
Db 3085 GCCAGTGTGAAGGAGGGGCTGAGGCGCTCCAGGACCTTGAAGGCGCAATCAACAGGTG 3144  
Qy 121 -----AlaAlaGluThrLeuProGluAla----- 128  
Db 3145 GGGACCCAGGCGCTCCAGGCTCCCGAGGGCTGCCAGAGGCGATAGAGCCCTGGTGGAA 3204  
Qy 128 ----- 128  
Db 3205 GATGATGTGCCCCAGGGGGTGACCAAGCCTCCACAGAGTCATGTTGGGGTCAGAGCCT 3264  
Qy 129 -----SerArgAlaValGluAsnGlyCysCysThrProLysGluGlyArg 143  
Db 3265 GCCATGGGTGAGTCTCTCGGGAGCTGAGCCAGGCGCTGGGGCAGGGGTGGGAGGCTG 3324  
Qy 144 GlyAlaProAlaGluAlaGlyLysGluGln-----LysGluThrAsn 157  
Db 3325 GGGGACCCAGGCGATCTGACAGGGAAGAGGTGATGGAACCAACCCCTGGAGAGGAGGT 3384  
Qy 158 IleGluSerMetLysMetGluGlySerArgLysArgLysLeuArg-----Gly 172  
Db 3385 TTGGAGGCAAGAGAGGCTTCCAGGCTTGGAAAGGGCTTGAAGAGGACCTAGAGAGGAGGT 3444  
Qy 173 GlyLeuGlyTrpGluSerSerLeuArgGlnArgProMetProArgLeuThrPheGlnAla 192  
Db 3445 GGTCTGGGACAGAGTCTCC-----GAGCTGCTGGGAGAGGC 3483  
Qy 193 GlyAspProTyrTyrIleSerLysArgLysArgAspGluTrpLeuAlaArgTrpLysArg 212  
Db 3484 AGAGACCTTTGGAGCGCTCCAGGAGGGGTAGGAGGAGTCAGAGGCTGAGGCCCCCAGG 3543  
Qy 213 GluAlaGluLys-----LysAlaLysValIleAlaGlyMetAsnAlaValGluGluAsn 230  
Db 3544 GGAGCAGAGGAGGCGTTCCTCTGTGAGACCTGGGCCACACTGGAAGTGTATGCCCTTCA 3603  
Qy 231 GlnGlyProGlyGluSerGlnLysValGluGluAlaSerProProAlaValGlnGlnPro 250  
Db 3604 CCTTGGCTCTGGGTTCAGAGGAGGAGTACACAGGATGCTGCTGCTGCTGCTGCTGCT 3663  
Qy 251 ThrAspProAlaSerProThrValAlaThrThrProGluProValGlySerAspAlaGly 270  
Db 3664 -----AGCCCAACGTACACC----- 3678  
Qy 271 AspLysAsnAlaThrLysAlaGlyAspGluProGluTyrGluAspGlyArgGlyPhe 290  
Db 3679 -----CCGATCTCTGGAAGATGCCCTGGGCTC 3705  
Qy 291 GlyIleGlyGluLeuValTrpGlyLysLeuArgGlyPheSerTrpTrpProGlyArgIle 310





QY 238 LysValGluGluAlaSerProProAlaValGln-----Gln 249  
Db 560 -----CCAGCCCTCCCGCTCGAGGTACACAGCTCTTGGTGGCTGTTGGC 607  
QY 250 ProThrAspProAlaSerProThrValAlaThrThrProGluProValGlySerAspAla 269  
Db 608 CCCGAACCTCCCGAAGATCCACCGGCTGCCCGCCACCGAGGGGTTGTCGCCGCTC 667  
QY 270 GlyAspLysAsnAlaThrLysAlaGlyAspAspGluProGluThrGluAspGlyArgGly 289  
Db 668 ATGAGCCGGTCGGGTGCAAGTGGAGACAGTCC----- 703  
QY 290 PheGlyIleGlyGluLeuValTrpGlyLysLeu-----ArgGlyPheSerTrpTrpPro 307  
Db 704 -----GGGACGCGACGTCGCCATAAAGTGTGCCCGGGGCGCTGTCA-----CCA 748  
QY 308 GlyArgIleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGluGlyThrArgTrp 327  
Db 749 GCCCGG----- 754  
QY 328 ValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLysLeuMetProLeu 347  
Db 755 -----CAGCTGCTGCTCCCGGCC 772  
QY 348 SerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnProMetTyrArgLys 367  
Db 773 TCT-----GAGAGCCCTCACTGTCGGGGGCCAGTGAAGCGTCTCCGAGGCC 823  
QY 368 AlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeuPheProValCys 387  
Db 824 CTTGGCGTGGAGTTGAGGAG----- 844  
QY 388 HisAspSerAspGluSerAspThrAlaLysAlaValGluValGlnAsnLysProMetIle 407  
Db 845 GAGATGCTCTGAGTCCGAGGAGTCTCGGGTCCGCTTCTGAAGGCGCAACCTCGG-- 901  
QY 408 GluTrpAlaLeuGlyGlyPheGlnProSerGly-----ProLysGly---- 421  
Db 902 -----GCTCTGGTGGCGCGCGGTGGAGGAGGAGCCCGGCTGTCGCCCGGGGGCG 955  
QY 422 -----LeuGluProProGluGluGluLysAsnProTyrLysGluVal 435  
Db 956 GCAGCAGGAGGCGTCGCCCTGTGTCGCCAAGGAAGATTCCCGCTTCTCAGCGCCAGGGTC 1015  
QY 436 TyrThrAspMetTrpValGluProGluAlaAlaAlaTyrAlaProProProAlaLys 455  
Db 1016 -----GCCCTGTGGAGCAGGAC-----GCGCGGATGGCGCCCGGGCGC 1054  
QY 456 LysProArgLysSerThrAlaGluLys-----ProLysValLys 468  
Db 1055 TCCCGCTGGCCACACGAGTGTGATGATTTTCATCCAGTGCCTATCTGCTCTCAATCAC 1114  
QY 469 GluIleAlaAspGluArgThrArgGluArgLeuValTyrGlu--ValArgGlnLysCysA 488  
Db 1115 GCCTATTGGCAGCCCGCACTGGCAGCTGCTGGAGACAGAAAGTTACGAGCGGGGCC 1174  
QY 488 rgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThrLeuGluHis-Pro 507  
Db 1175 GGGGTGCGCAGCGCTTTCG-----CCGCGCGGAGTTTCACCTGTGCTGTCTCCACCCCG 1231  
QY 508 LeuPheValGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGluCysAlaTyrGln 527  
Db 1232 GTCGCTGTAGGCGAC-----TTCCCGAGTGGCGGTACCG 1267  
QY 528 TyrAsp-----AspAspGlyTyrGlnSerTyrCysThrIle----- 539  
Db 1268 CCCGACCGGAGCCCAAGGACGACGCTATCCCTCTCTATAGCGAGTTCCAGCCCGCGCT 1327  
QY 540 -----CysCysGlyGlyArgGluVal 546  
Db 1328 CTAAGATAAAGGAGGAGGAGGAGGCGGAGGCGCTCCCGCGCTCCCGCGTTCCTAC 1387

QY 547 LeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysValAspLeuLeu 566  
Db 1388 CTTGTGGCGGTGCCAAC----- 1405  
QY 567 ValGlyProGlyAlaAlaGlnAlaAlaIleLysGluAspPro-TrpAsnCysTyrMetCys 586  
Db 1406 -----CCCGCAGCCTTCCCGATTTCCCGTTGG----- 1433  
QY 586 sGlyHisLysGlyThrTyrGlyLeuLeuArgArgGluAspTrpProSerArgLeuGln 606  
Db 1434 -GCCACCGCCCGCCG-----TGCCCGCGCAGGACCCCATCCA----- 1472  
QY 606 nMetPhePheAlaAsnHisAspGlnGluPheAspProProLys-----Va 622  
Db 1473 -----GACCCGGGAAGCGCGGTGACGGCC 1498  
QY 622 lTyrProProValProAlaGluLysArgLysProIleArg 635  
Db 1499 GCACCGCGCAGTGCCTCAGTCTCGTCTGCTCTCTCTCGG 1538

RESULT 7  
US-08-759-873-1  
; Sequence 1, Application US/08759873  
; Patent No. 5683885  
; GENERAL INFORMATION:  
; APPLICANT: Kieback, Dirk G.  
; TITLE OF INVENTION: METHODS FOR DIAGNOSING AN INCREASED RISK  
; TITLE OF INVENTION: OF BREAST OR OVARIAN CANCER  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SUGHRUE, MIOM, ZINN, MACPEAK & SEAS  
; STREET: 2100 Pennsylvania Avenue, N.W., Suite 800  
; CITY: Washington, D.C.  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/759,873  
; FILING DATE: 12-APRIL-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: KIT, Gordon  
; REGISTRATION NUMBER: 30,764  
; REFERENCE/DOCKET NUMBER: A-6612  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 293-7060  
; TELEFAX: (202) 293-7860  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3014 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
US-08-759-873-1

Alignment Scores:  
Pred. No.: 0.0046 Length: 3014  
Score: 166.00 Matches: 146  
Percent Similarity: 31.21% Conservative: 65  
Best Local Similarity: 21.60% Mismatches: 209  
Query Match: 3.36% Indels: 257  
DB: 1 Gaps: 31

US-09-720-086-7 (1-912) x US-08-759-873-1 (1-3014)

QY 53 ArgLysArgLysHisProProValGluSerGlyAspThrProLysAspProAla----- 70

Db	9	CGCGCCCTCCCCCGCCCTGACCCAGGA-gtggagatccctccggatccagccacattc	67
QY	71	-----valiserlyserProsermetalaglasp	81
Db	68	AACACCACTTCTCTCTCCCTCTGCCCCCTATATTCCGAAACCCCTCTCTCTCCCTTT	127
QY	82	SerGlyAlaSerGluLeuLeuPro--AsnGlyAspLeuGluLysArgSer-----	97
Db	128	TCC-----CTCTCTCTGAGACGGGGAGAGAAAAAGGGAGTCCACTGTC	175
QY	98	-----GluProGln	100
Db	176	ATGACTAGCTGAAGCAAAAGGCTCCCCGGGCTCCCCAGCTGGCGGGCGGCCCTCTCC	235
QY	101	ProGluGluGlySerProAlaGlyGlyGlnLysGlyAlaProAlaGluGlyGluGly	120
Db	36	CCGAGGTCGGAATCCCACTGCTGTGTGCGCCAGCCGAGGTCCTCCGGGGAGCCAG	295
QY	121	AlaAlaGluThrLeuProGluAlaSerArgAlaValGlu-----AsnGlyCysCys	137
Db	296	ACCTCGGACACTTGTCTGAAGTTTC--GCCATACCTATCTCCCTGAGGGCTACTC	352
QY	138	ThrProLysGluGlyArgGlyAlaProAlaGluAlaGlyLysGluGlnLysGluThrAsn	157
Db	353	TTCCTCTGGCCCTGCCAGGACAGGACCCCTCCGAGAAAGACGACGACGACGACTG	412
QY	158	IleGluSerMetLysMetGluGlySerArgGlyArgLeuArgGlyGlyLeuGlyTyrGlu	177
Db	413	CTGCGGACGAGGAGGGCGCATATTCCAGAGCTGAAGCTAACAGGGGTGCTGAGAGGCAGC	472
QY	178	SerSerLeuArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTyrTyr	197
Db	473	AGTTCT-----	478
QY	198	IleSerLysArgLysArgAspGluTyrLeuAlaArgTyrLysArgGluAlaGluLysLys	217
Db	479	--AGTCCCCAGAAAAAGACAGCGACTG-----	505
QY	218	AlaLysValIleAlaGlyMetAsnAlaValGluGluAsnGlnGlyProGluGluSerGln	237
Db	506	-----CTGACAGTGTTCTTGACACTCTGTTGGCGCCTCAGAGTCCCGGGCAGAGCCA	559
QY	238	LysValGluGluAlaSerProProAlaValGln-----Gln	249
Db	560	-----CCGAGCCCTCCCGCTGCGAGGTCAACAGCTCTGTGTCCTGTTGGC	607
QY	250	ProThrAspProAlaSerProThrValAlaThrThrProGluProValGlySerAspAla	269
Db	608	CCCGAACTTCCCGAAGATCCACCGGCTGCCCGGCCACCCAGGGGGTGTTCGCCGCTC	667
QY	270	GlyAspLysAsnAlaThrLysAlaGlyAspAspGluProGluTyrGluAspGlyArgGly	289
Db	668	ATGAGCGGATCGGGGTCAAGGTTGGAGACGCTC-----	703
QY	290	PheGlyIleGlyGluLeuValTyrGlyLysLeu-----ArgGlyPheSerTyrTyrPro	307
Db	704	-----GGACGGCAGCTGCCCATTAAGTGTGCCCGGGGCTGTCA-----CCA	748
QY	308	GlyArgIleValSerTyrTyrMetThrGlyArgSerArgAlaAlaGluGlyThrArgTyr	327
Db	749	GCCCGG-----	754
QY	328	ValMetTyrPheGlyAspGlyLysPheSerValValCysValGluLysLeuMetProLeu	347
Db	755	-----CAGCTGCTCTCCGGCC	772
QY	348	SerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnProMetLysArgLys	367
Db	773	TCT-----GAGAGCCCTCATGTGTCGGGGGCCCAAGTGAAGCCGTCTCCGAGGCC	823
QY	368	AlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeuPheProValCys	387

```

Db      824 GCTCGGATGGAGTTGAAGAC-----                        844
QY      388 HisAspSerAspIuSerAspThrAlaLysAlaValGluValGlnInenLyPProMetIle 407
          |||         |||||:::   ::|         :|||
Db      845 GAGATAGGTCTGAGTCCGAGAAGACTGCCGGTCCTCCTGAAGGCMAAACCCTCG-- 901
QY      408 GluTrpAlaLeuGlyGlyPheGlnProSerGly-----ProLysGly-- 421
          |||||       |||||::|  |||         |||
Db      902 -----GCTCTGGGTGGCCCGCGCGCTGAGAGAGAGCCCGCGCTGCCCGGGGGCG 955
QY      422 -----LeuGluProProGluGluGluLysAsnProTyrlLysGluVal 435
          |||         |||         |||         |||
Db      956 GCAGAGAGAGGCGTCCGCTCGTGTCCCAAGAAATTCCCGCTTCAGCGCCCAAGGTC 1015
QY      436 TyrThrAspMetTrpValGluProGluAlaAlaAlaTyrAlaProProProAlaLys 455
          |||||       |||||::|  |||         |||
Db      1016 -----GCCCTGGTGAGCACAGAC-----GCCCGAATGGCGCCCGGGCGC 1054
QY      456 LysProArgLysSerThrAlaGluLys-----ProLysValLys 468
          |||         |||         |||         |||
Db      1055 TCCCCCGTCGGCCACCAACGCTGATGATTTCATCAGCGCTTCTCCTCATATCAC 1114
QY      469 GluIleIleAspGluArgThrArgGluArgLysLeuValTyrGlu--ValArgInLysCysA 488
          :|||:::   :|||:::   |||         |||
Db      1115 GCCTTATTGGCAACCCGCACTCGGCAGCTCTGGAAGACGAAGTTACAGCGCGGGGCC 1174
QY      488 rGAsnIlleGluAspIleCysIleSerCysGlySerLeuAsnValThrLeuGluHis-Pro 507
          :|||:::   :|||:::   |||         |||
Db      1175 GGGGCTGCCAGCGCCTTTGC---CCGCGCGGAGTTACCTCTGCTGCTCCACCCCG 1231
QY      508 LeuPheValGIyGlyMetCysGlnAsnCysLysAsnCysPheLeuGluCysAlaTyrGln 527
          :|||:::   :|||:::   |||         |||
Db      1232 GTCGGTGAGCGCAC-----TTCCCAGACTGGCGATACCG 1266
QY      528 TyrAsp-----AspAspGlyTyrGlnSerTyrCysThrIle----- 539
          |||         |||         |||         |||
Db      1268 CCCGACGCCGAGCCCAAGACGACGCTACCTCTATAGCACTTCACGCCCGCGCT 1322
QY      540 -----CysCysGlyIyArgGluVal 546
          |||         |||         |||         |||
Db      1328 CTAAGAATAAAGAGAGAGAGAGAGCGCGGAGCTCCGCGGCTCCCGCTTCTTAC 1387
QY      547 LeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysValAspLeuLeu 566
          |||         |||         |||         |||
Db      1388 CTGTGGCGCGGTGCCAAC-----                    1405
QY      567 ValGlyProGlyAlaAlaGlnAlaAlaIleLysGluAspPro-TripAsnCysTyrMetCy 586
          |||         |||         |||         |||
Db      1406 -----CCGCAAGCTTCCCGGATTTCCCGTTGG----- 1433
QY      586 sGlVHISLYSGlyThrTyrGlyLeuLeuArgArgArgGluAspTrpProSerArgLeuGI 606
          |||||       |||||::|  |||         |||
Db      1434 -GGCCACCGCCCGCGC-----TGCCGCGCGAGAGCAACCCATCCA- 1471
QY      606 mMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProLys-----Va 622
          |||||       |||||::|  |||         |||
Db      1473 -----GACCCGGGAAGCGCGGTGACGCGC 1495
QY      622 ITyrProProValProAlaGluLysArgLysProIleArg 635
          :|||:::   :|||:::   |||         |||
Db      1499 GCAACCGCAAGTGCCTCAGTCTGTGTCGCTCTCTCGG 1538

RESULT 8
US-09-085-199B-6
; Sequence 6, Application US/09085199B
; Patent No. 6235879
; GENERAL INFORMATION:
; APPLICANT: Hayden, Michael R.
; APPLICANT: Hackam, Abigail
; APPLICANT: Hug, A.H.M. Mahububul
; APPLICANT: Chopra, Vikramjit Singh
; APPLICANT: Kalchman, Michael
; TITLE OF INVENTION: Apoptosis Modulators That Interact with the

```









QY 232 ----- 232  
Db 2915 GAATAGATCAATTGACATGAAATTTGGGGTTCCTAATTTCTCTATGTAATTCGCAAGT 2856  
QY 233 -----ProGly 234  
Db 2855 CTGCTGTCAATCTCTCTGACTTTTCCATCCAAATCTCCGGGACCACTTCTCTACCCAG 2796  
QY 235 GluSerGlnLysValGluAlaSerPro---ProAlaValGlnGlnProThrAspPro 253  
Db 2795 AACTCCCGCCGCGCCAGCCCACTGCGCCCACTCCCACTCCCACTCCCACTCCCA 2736  
QY 254 AlaSerProThrValAlaThrThrProGluPro-----ValGlySerAspAlaGly 270  
Db 2735 CACTCCCGCCATCAGTGCACGCGCCCGGAGACTCCACATTTCTTGGGGGTCCCGAGGA 2676  
QY 271 AspLysAsnAlaThrLysAlaGlyAspAspGluProGluTyrGluAspGlyArgGlyPhe 290  
Db 2675 -----GACGGGCAAAAGATGA 2661  
QY 291 GlyIleGlyGluLeuValTrpGlyLysLeuArgGlyPheSerTrpTrpProGlyVargile 310  
Db 2660 TCCCTAGGTGCTCGGTGGGG-----GGTCCCAAGATCTTCTTCCAGG--- 2613  
QY 311 ValSerTrpTrpMetThrGlyArgSerArgAlaAlaGluGlyThrArgTrpValMetTrp 330  
Db 2612 -----AGCACACCGAGGAGTCAAGCTACTCTGTGG----- 2583  
QY 331 PheGlyAspGlyLysPhe-----SerValValCysValGluLysLeuMetPro 346  
Db 2582 -----GGCATGTCTGTCTAGCTACCGGCATCTCCAATGTGGCCAACTCTCTTCT 2532  
QY 347 LeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnProMetTyrArg 366  
Db 2531 CACTCTGTCTGGCCCAACACA----- 2511  
QY 367 LysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeuPheProVal 386  
Db 2510 -----CAGTAAGTAAGGTGTATCGGAATGGGTGGGTATGGGAGCATCTG 2463  
QY 387 CysHisAspSerAspGluSerAspThrAlaLysAlaValGluValGlnAsnLysProMet 406  
Db 2462 GCTCAGCTCGGGGGCGCTGGGACAGCGCCAGAGCGCCAGCTCACAAGGGCATCT 2403  
QY 407 IleGluTrpAlaLeuGlyGlyPheGlnPro---SerGlyProLysGlyLeuGluProPro 425  
Db 2402 CTGTCTGTGTGAAAGATGGAATTAGGCTTGGGACACACCCCGCCACCTACGGGTGTAT 2343  
QY 426 GluGluGluLysAsnProTyrLysGluValTyrThrAspMetTrpValGluProGluAla 445  
Db 2342 CCCAAATGCGCGCGCCACACTCAG----- 2316  
QY 446 AlaIleTyrAlaProProProProAlaLys-----LysProArgLysSer 460  
Db 2315 GCCGACCATCTGAGCCCTCTCTCTGAGAGCGGCGACAGACAGCGCGCGAGGGCAGT 2256  
QY 461 -----ThrAlaGluLysProLysValLysGlu-IleIleAspGluAr 474  
Db 2255 GGAGATGGCAACAGAGGGGGCTCAGACCATCAAAACATTTGGAGATCTTAGCGGACT 2196  
QY 474 gThrArgGluArgLeuValTyrGluValArgGlnLysCysArgAsnIleGluAspIleCy 494  
Db 2195 TTGGCCTGATCATCAGATAT-----CGTAGTTGCTTAGATATGCTTATTGTTT 2145  
QY 494 sile---SerCysGlySerLeuAsnValThrLeuGluHisProLeuPheValGlyGly-- 512  
Db 2144 CATCTAATTTTGGGGGTGGTGTGTGTGCGCATGTGTGTGCGAGGTGGGTGCTTG 2085  
QY 513 -MetCysGlnAsnCysLysAspPheLeuGluCysAlaTyrClnTyrAspAspG1 532  
Db 2084 CGTGTGCAACGTGTGTGCTGCGGTGTTGTGTGTGCTGAGCTGTGTCTATGTATGTGCTGC 2025

QY 532 yTyrGlnSerTyrCysThrIleCysCysGlyGlyArgGluValLeuMetCysGlyAsnAs 552  
Db 2024 GCATGAGTCA 1965  
QY 552 naenCysAspGlyCysPheCysValGluCysValAspLeuValGlyProGlyAlaAl 572  
Db 1964 TGTGTCTGT 1916  
QY 572 aGlnAlaIleLysGluAspProTyrAsnCysTyrMetCysGlyHis 588  
Db 1915 -----GTGTTTCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1889

RESULT 11  
US-08-949-386-23  
; Sequence 23, Application US/08949386  
; Patent No. 6090623  
; GENERAL INFORMATION:  
; APPLICANT: Harpold, Michael  
; APPLICANT: Ellis, Steven  
; APPLICANT: Williams, Mark  
; APPLICANT: McCue, Ann  
; APPLICANT: Gillespie, Alison  
; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 38  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Martin, Haller & McClain  
; STREET: 1650 Union Street  
; CITY: San Diego  
; STATE: California  
; COUNTRY: US  
; ZIP: 92101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/949,386  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/290,012  
; FILING DATE: 11-AUG-1994  
; APPLICATION NUMBER: 08/149,097  
; FILING DATE: 5-NOV-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/105,536  
; FILING DATE: 11-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie L.  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 519808  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 238-0999  
; TELEFAX: (619) 238-0062  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7791 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 237..7037  
; OTHER INFORMATION: /standard\_name="Alpha-1A-2"  
US-08-949-386-23

Alignment Scores:  
Pred. No.: 0.0401 Length: 7791  
Score: 161.00 Matches: 165  
Percent Similarity: 29.60% Conservative: 73



Best Local Similarity: 20.52% Mismatches: 234  
 Query Match: 3.26% Indels: 334  
 DB: 3 Gaps: 41  
 US-09-720-086-7 (1-912) x US-08-949-386-23 (1-7791)

OY 7 SerGlyProGlyAspThr-----SerSerSerAlaAlaGluArgGluGlu 21  
 Db 2636 AATGACCCCGACGACGCGCTGGAAGGCTGCTACACGCGGACCTGCGGACGACATGAA 2695  
 OY 22 Asp-----ArgLysAspGlyGluGluGluGluProArgGlyLysGlu 37  
 Db 2696 GACGCACTTGACCGCGCGCTGTGTGTGACCGCAGAGAACCGCAA---CAACAACAC 2752  
 OY 38 ArgGluGluProSerThrThrAlaArgLys-----ValGlyArgProGlyArgLysArg 55  
 Db 3753 CAACAAAGACCG 2812  
 OY 56 LysHisProProValGluSerGlyAspThrProLysAspProAlaValIleSerLysSer 75  
 Db 2813 GAACCTTCTCAG---GAAACACGCGCGCTACACACATCG-----GCGCGGGA 2857  
 OY 76 ProSerMetAlaGlnAspSerGlyAlaSerGluLeu-----ProAsnGlyAsp 92  
 Db 2858 CCCGACGCGCTCGCGCGCGCTGACGACGAGCGCGCTGCGCGGGAACGACGAGCGCGA 2917  
 OY 93 LeuGluLysArgSerGluProGluProGluGluGlySerProAlaGlyGlyGluLysGly 112  
 Db 2918 GCTGACGCGGAGGAGGACCTTACGCGCGCGAGTGCACCA-----CGCGCGGGA 2968  
 OY 113 GlyAlaProAlaGluGlyGluGlyAlaAlaGluThrLeuProGluAlaSerAlaVal 132  
 Db 2969 GGGGAGCGCT-----GGAGCAACCGCGGCTTCTGGAGGCGGAGCGCGAGCG--- 3013  
 OY 133 GluAsnGlyCysCysThrProLysGluGlyArgGlyAlaProAlaGlu----- 148  
 Db 3014 -----AGGCAAGCGCGGAGACCCCGACCGAGGACACGTGACCG 3052  
 OY 149 AlaGlyLysGluGluLysGluThrAsn----- 157  
 Db 3053 GCAGGGGAGGACGAGGAGGACCGGACGCGGCTCCCGACGAGGCGCGGACGAGGACGA 3112  
 OY 158 IleGluSerMetLysMetGluGlySerArgGlyArgLeuArgGlyLysGlyTyrGlu 177  
 Db 3113 TCGACGTATCGCGCGGACCGGACCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3172  
 OY 178 SerSerLeuArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTyrTyr 197  
 Db 3173 GAGGCGCGGACCGGCA-----GGGACGCGCGCGCGC 3205  
 OY 198 IleSerLysArgLysArgAspGluTyrPheAlaArgTyrLysArgGluAlaGluLys 217  
 Db 3206 CCGGCGCGCGGAGGCGGAGGCGGAGGCGCGGCGCGGCGGCGGCGGCGGCGGCGG 3259  
 OY 218 AlaLysVal-----IleAlaGlyMetAsnAlaValGluGluAsnGlnGly 232  
 Db 3260 GCACCGCGCATGCGCTCCAGCACCGATACGAGGGGAGCGCG---GAGGAGGACAAAGA 3316  
 OY 233 ProGlyGluSerGlnLysValGluGluAlaSerProProAlaValGlnGlnProThrAsp 252  
 Db 3317 -----GCGGAGGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3361  
 OY 253 ProLaserProThrValAlaThrThrProGluProValGlySerAspAlaGlyAspLys 272  
 Db 3362 -----TGTGTGCGGCGCGCAACCTGTC-----AACCAAC 3388  
 OY 273 AsnAlaThrLysAlaGlyAspAspGluProGluTyrGluAspGlyArgGlyPheGlyLe 292  
 Db 3389 CCGGCAATTCACAGACGCTGGGCGGCAAGACCCCTGCGAGAGATATGACAA 3448  
 OY 293 GlyGluLeuValTyrGlyLysLeuArgGlyPheSerTyr-----TyrPro 307  
 Db 3449 CATGAAGAACAAACAGCTGGCGACCGCGAGTGGCGCGCTCCCGACGCGACCTTGGCCA 3508

OY 308 GlyArgIleValSerTyrTrpMetThrGlyArgSerArgAlaAlaGluGlyThrArgTyr 327  
 Db 3509 -----CGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3535  
 OY 328 ValMetTyrPheGlyAspGlyLysPheSerValValCysValGluLysLeuMetProLeu 347  
 Db 3536 -----GATGGGAA----- 3544  
 OY 348 SerSerPheCysSerAlaPheIleGlnAlaThrTyrAsnLysGlnProMetTyrArgLys 367  
 Db 3545 -----CAGCACCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3562  
 OY 368 AlaIleTyrGluValLeuGluValAlaSerSerArgAlaGlyLysLeuPheProValCys 387  
 Db 3563 -----CATGTGGCCAT-----CCCTGC 3580  
 OY 388 HisAspSerAspGluSerAspThrAlaLysAlaValGluValGlnAsnLysProMetIle 407  
 Db 3581 CATGGCCACCAACCCCGACAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3625  
 OY 408 GluTyrAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGluProProGluGlu 427  
 Db 3626 -----CAACCGGGAACCATCCATCCATCCGCGCGCGCGCGCGCGCGCGCGCG 3664  
 OY 428 GluLysAsn---ProTyrLys-----GluValTyrThrAspMetTyr----- 440  
 Db 3665 CCCCGAGATAGCTTATCTGTACACCAACCCCGCGGACCGGACCGGACCAATTCAGCTAGAC 3724  
 OY 441 -----ValGluProGluAlaAlaAlaTyrAlaPro-----ProProProAlaLys 455  
 Db 3725 TGCCAGGAACCGGACACACACAGTGGACATCCCCCGCGCGCGCGCGCGCGCGCGCG 3784  
 OY 456 LysProArgLysSerThrAlaGluLysProLysValLysGluIleLeuAspGluArgThr 475  
 Db 3785 CCACACCGCTGTACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 3844  
 OY 476 ArgGluArgLeuValTyrGluValArgGluLysCysArgAsnIleGluAspIleCysIle 495  
 Db 3845 AGAGAAAGAA----- 3853  
 OY 496 SerCysGlySerLeuAsnValThrLeuGluHisProLeuPheValGlyMetCysGln 515  
 Db 3854 -----GGAGGA----- 3859  
 OY 516 AsnCysLysAsnCysPheLeuGluCysAlaTyrGlnTyrAspAspGlyTyrGlnSer 535  
 Db 3859 ----- 3859  
 OY 536 TyrCysThrIleCysCysGlyGlyArgGluValLeuMetCysGlyAsnAsnAsnCysCys 555  
 Db 3860 -----GAGGAGAGACGA----- 3871  
 OY 556 ArgCysPheCysValGluCysValAspLeuLeuValGlyProGlyAlaAlaGlnAla 575  
 Db 3871 ----- 3871  
 OY 576 IleLysGluAspProThrAsnCysTyrMetCysGlyHisLysGlyThrTyrGlyLeuLeu 595  
 Db 3872 -----CCGTGGGA----- 3880  
 OY 596 ArgArgArgGluAspTyrProSerArgLeuGlnMetPhePheAlaAsnAsnHisAspGln 615  
 Db 3881 ---AGACGCGCTTACGCAATGCTCTCTATAGCTCATGTTATCTGTCCACGACCAA 3937  
 OY 616 GluPheAspProProLysValTyrProProValProAla----- 628  
 Db 3938 CCCCTTGGCGCGCTGTGCATTCATCTGAACCTGCGCTTGTGAGATGTCATCTCT 3997  
 OY 629 -----GluLysArgLysProIleArg----- 635  
 Db 3998 CATGTGATTCGCATAGCAGCATGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4057



```
QY 636 -----ValLeuSerLeuPheApeGlyIleAlaThrGlyLeuValLeuLy 651
Db 4058 ACCTCGGAACACGCTCGGATACCTTTTACAGGCTCTTCACTTT-- 4115
QY 651 sAspLeuGlyIleGlnValAspArgTyrIleAlaSerGluValCysGluAspSerIleTh 671
Db 4116 -----GAGATGCTGATCAAGATGATTGA 4138
QY 671 rValGlyMetValArgHisGlnGlyLysIleMetTyrValGlyAspValArgSerValTh 691
Db 4139 CCTGGGCTCTGCTCGCATCAGGCT-----GCCTACTTCCGTGACCTCTGGAATATTCT 4192
QY 691 rGlnLysHisIleGlnGluTrpGlyProPheAspLeuValIleGlyGlySerProCysAs 711
Db 4193 -CGACTTCATAGTGTGCTCAGTGGGCCCT-----GGTAGCCTTTGCT 4233
QY 711 nAspLeuSerIle-----ValAspProAlaArg---LysG1 722
Db 4234 -TCACTGGCAATAGCAAGAAAGACATCAACACGATTAAATCCCTCCGAGTCTCTCCGG 4292
QY 722 YLeuTyrGlu 725
Db 4293 GTGCTACGAC 4302

RESULT 12
US-08-450-562-23
; Sequence 23, Application US/08450562
; Patent No. 6098514
; GENERAL INFORMATION:
; APPLICANT: Harpold, Michael
; APPLICANT: Ellis, Steven
; APPLICANT: Williams, Mark
; APPLICANT: McCue, Ann
; APPLICANT: Gillespie, Alison
; APPLICANT: Feldman, Daniel
; APPLICANT: Brenner, Robert
; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; CITY: San Diego
; STATE: California
; COUNTRY: US
; ZIP: 92101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,562
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/404,950
; FILING DATE: 13-MAR-1995
; APPLICATION NUMBER: 08/336,257
; FILING DATE: 7-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/314,083
; FILING DATE: 28-SEPT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/311,363
; FILING DATE: 23-SEPT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290,012
; FILING DATE: 11-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/223,305
; FILING DATE: 4-APR-1994
; PRIOR APPLICATION DATA:
```

```
; APPLICATION NUMBER: 08/193,078
; FILING DATE: 07-FEB-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/149,097
; FILING DATE: 5-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/105,536
; FILING DATE: 11-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/914,231
; FILING DATE: 13-JULY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/868,354
; FILING DATE: 10-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06903
; FILING DATE: 14-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/745,206
; FILING DATE: 15-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/620,250
; FILING DATE: 30-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/603,751
; FILING DATE: 08-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/482,384
; FILING DATE: 02-FEB-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US89/01408
; FILING DATE: 04-APR-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/176,899
; FILING DATE: 04-APR-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-519812
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 238-0999
; TELEFAX: (619) 238-0062
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7791 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 237..7037
; OTHER INFORMATION: /standard_name= "Alpha-1A-2"
US-08-450-562-23

Alignment Scores:
Pred. No.: 0.0401 Length: 7791
Score: 161.00 Matches: 165
Percent Similarity: 29.60% Conservative: 73
Best Local Similarity: 20.52% Mismatches: 234
Query Match: 3.26% Indels: 334
DB: 3 Gaps: 41

US-09-720-086-7 (1-912) x US-08-450-562-23 (1-7791)
QY 7 SerGlyProGlyAspThr-----SerSerSerAlaAlaGluArgGluGlu 21
Db 2636 AATGGACCCGGACGAGCGCTGGAAGGCTGCCTACACGCGGCGCTGCGCCAGACATGAA 2695
QY 22 Asp-----ArgLysAspGlyGluGluGluGluProArgGlyLysGluGlu 37
Db 2696 GACGCACTTGGACCGCGCTGTGTGGTGGACCGGAGAGACCGCAA---CAACACAC 2752
```

QY 38 ArgGlnGluProSerThrThrAlaArgLys-----ValGlyArgProGlyArgLysArg 55  
Db 2753 CAAAGAGAGCCGGGGCCGAGCCGACCGTGGACCAAGCCGCTCGGCGACAGCGCGCGCA 2812  
QY 56 LysHisProProValGluSerGlyAspThrProLysAspProAlaValIleSerLysSer 75  
Db 2813 GGACTTCTCTAC--GAAACAGGCGCCGCTTACCAAGATCG-----GGCCCGGGA 2857  
QY 76 ProSerMetAlaGlnAspSerGlyAlaSerGluLeu-----ProAsnGlyAsp 92  
Db 2858 CCCCAGCGGCTCGCGGGCTGGAGCGACGAGAGCCCTGGCGGGGAAGCCAGAGCGCGCA 2917  
QY 93 LeuGluLysArgSerGluProGlnProGluGluGlySerProAlaGlyGlyGlnLysGly 112  
Db 2918 GCTGACCGGGAGGAGGACCTTACGCGCGCACTGGACCACCA-----CGCCCGGA 2968  
QY 113 GlyAlaProAlaGluGluGlyGluGlyValAlaGluThrLeuProGluAlaSerArgAlaVal 132  
Db 2969 GGGGCAAGCT-----GGAGCAACCCGGGTTTGGAGGGCGAGGCCGAGCG-- 3013  
QY 133 GluAsnGlyCysCysThrProLysGluGlyArgGlyValAlaProAlaGlu----- 148  
Db 3014 -----AGGCAAGGCGGGGACCCCGACCGAGGACCGTGCACCG 3052  
QY 149 AlaGlyLysGluGlnLysGluThrAsn----- 157  
Db 3053 GCAGGGGGGCGACGAGAGAGCCGACCGGGTCCCGCGACGGGCGCGGACCGGAGCA 3112  
QY 158 IleGluSerMetLysMetGluGlySerArgGlyArgLeuArgGlyGlyLeuGlyTrpGlu 177  
Db 3113 TCGACGTATCGCGCGACCGACCGCGCGGGAGAGGTCGCGAGAGCAAGCGCGAGCG 3172  
QY 178 SerSerLeuArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTyr 197  
Db 3173 GAGGCGCGCGACCGCA-----GGCGAGCGCGCGCGC 3205  
QY 198 IleSerLysArgLysArgAspGluTrpLeuAlaArgTrpLysArgGluAlaGluLysLys 217  
Db 3206 CCGGGGCGCGAGGCGAGGCGCGAGGCGCGCGCGCGCGCGA-----CGCGAGAGAG 3259  
QY 218 AlaLysVal-----IleAlaGlyMetAsnAlaValGluGluAsnGlnGly 232  
Db 3260 GCACCGGCGTGGCGCTCCAGCCACGTACGAGGGGAGCGCGC--GAGGAGGACAAAGA 3316  
QY 233 ProGluLysSerGlnLysValGluGluAlaSerProProAlaValGlnGlnProThrAsp 252  
Db 3317 -----GCGAGGCGATCGGAGAGAGAAAGAACACCAAGGCTCGGGGTCC----- 3361  
QY 253 ProAlaSerProThrValAlaThrThrProGluProValGlySerAspAlaGlyAspLys 272  
Db 3362 -----TGTGTCCGGCCCAACTGTC-----AACCAAC 3388  
QY 273 AsnAlaThrLysAlaGlyAspAspGluProGluTrpGluAspGlyArgGlyPheGlyIle 292  
Db 3389 CCGGCGAATCCAGACGACCTGGCGCGCGCAAGACCCACCCCTGGCAGAGATATTGACAA 3448  
QY 293 GlyGluLeuValTrpGlyLysLeuArgGlyPheSerTrp-----TrpPro 307  
Db 3449 CATGAAGAACACAGCTGGCCACCGCGAGTGGCGCGCTCCCGACGGCAGCGCTTGGCA 3508  
QY 308 GlyArgIleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGluGlyThrArgTrp 327  
Db 3509 -----CGCCGGCTCGCCCGACGAGCCAGCCAGCA--- 3535  
QY 328 ValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLysLeuMetProLeu 347  
Db 3536 -----GATGGGAAA----- 3544  
QY 348 SerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnProMetTyrArgLys 367  
Db 3545 -----CAGCACCGACCGCGCGC----- 3562  
QY 368 AlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeuPheProValCys 387

Db 3563 -----CATGCTGGCAAT-----CCCTGC 3580  
QY 388 HisAspSerAspGluSerAspThrAlaLysAlaValGlnValGlnAsnLysProMetIle 407  
Db 3581 CATGGCCACCAACCCCGAGAACCGCGCGCGCGCGCA----- 3625  
QY 408 GluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGluProProGluGlu 427  
Db 3626 -----CAACCGGGGAACCATTCGAATCCGCGCGCGCGCAAGAC 3664  
QY 428 GluLysAsn---ProTyrLys-----GluValTyrThrAspMetTrp----- 440  
Db 3665 CCCCAGAAATAGCCTTATCTCATCCCAACCCGCGGACCCAGACCAATTCAGCTAAGAC 3724  
QY 441 -----ValGluProGluAlaAlaAlaTyrAlaPro-----ProProAlaLys 455  
Db 3725 TGCAGAGAAACCCGACCAACCAACAGTGAACATCCCGCGAGCTGCCACCCCTCCAA 3784  
QY 456 LysProArgLysSerThrAlaGluLysProLysValLysGluIleLeuAspGluArgThr 475  
Db 3785 CCACACCGCTCGTACAGTGAACAAACCAACCCAGACCCACTGCCAAAAAAGAGGA 3844  
QY 476 ArgGluArgLeuValTyrGluValArgGlnLysCysArgAsnIleGluAspIleCysIle 495  
Db 3845 AGAGAAAGAA----- 3853  
QY 496 SerCysGlySerLeuAsnValThrLeuGlnHisProLeuPheValGlyGlyMetCysGln 515  
Db 3854 -----GAGGA----- 3859  
QY 516 AsnCysLysAsnCysPheLeuGluCysAlaTyrGlnTyrAspAspGlyTyrGlnSer 535  
Db 3859 ----- 3859  
QY 536 TyrCysThrIleCysCysGlyGlyArgGluValLeuMetCysGlyAsnAsnAsnCysCys 555  
Db 3860 -----GAGAGAAAGCA----- 3871  
QY 556 ArgCysPheCysValGluCysValAspLeuLeuValGlyProGlyAlaAlaGlnAlaAla 575  
Db 3871 ----- 3871  
QY 576 IleLysGluAspProTrpAsnCysTyrMetCysGlyHisLysGlyThrTyrGlyLeuLeu 595  
Db 3872 -----CCGTGGGGA----- 3880  
QY 596 ArgArgArgGluAspTrpProSerArgLeuGlnMetPhePheAlaAsnAsnHisAspGln 615  
Db 3881 ---AGACGGCCCTAAGCAATGCTCTCATAGCTCATGTTCACTCTCCACAGACCAA 3937  
QY 616 GluPheAspProProLysValTyrProProValProAla----- 628  
Db 3938 CCCCCTGCGCGCTGAGCATTCATCATCTGAACTCGCTACTTGAAGTGCATCCT 3997  
QY 629 -----GluLysArgLysProIleArg----- 635  
Db 3998 CATGGTATTGCATGACAGCATCGCCTGCGCGAGAGACCTGTGCAGCCCAAGC 4057  
QY 636 -----ValLeuSerLeuPheAspGlyIleAlaThrGlyLeuLeuValLeuLys 651  
Db 4058 ACCTCGAACAACAGCTGCGATCTTGACTTGAAGCTTTTACAGCGCTTCACTTT-- 4115  
QY 651 AspLeuGlyIleGlnValAspArgTyrIleAlaSerGluValCysGluAspSerIleTh 671  
Db 4116 -----GAGATGGATCAAGATGATGA 4138  
QY 671 rValGlyMetValArgHisGlnGlyLysIleMetTyrValGlyAspValArgSerValTh 691  
Db 4139 CCTGGGCTGCTGCTGATCAGGGT-----GCCTACTTCGCTGCTGATATATCT 4192  
QY 691 rGlnLysHisIleGlnGluTrpGlyProPheAspLeuValIleGlyGlySerProCysAs 711

Db 4193 -CGACTTCATAGTGTGTCAGTGGGGCCCT-----GGTAGCCTTTGGCT 4233  
Qy 711 nAspleuSerIle-----ValAsnProAlaArg---LysG1 722  
Db 4234 -TCATGGCAATAGCAAGAAAGAGATCAACACACGATTAAATCCCTCCGAGTCTCCGG 4292  
Qy 722 yLeuTyrGlu 725  
Db 4293 GTGCTACGAC 4302

RESULT 13  
US-08-984-709A-23  
; Sequence 23, Application US/08984709A  
; Patent No. 6320032  
; GENERAL INFORMATION:  
; APPLICANT: Williams, Mark E.  
; APPLICANT: Stauderman, Kenneth A.  
; APPLICANT: Harpold, Michael M.  
; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND  
; TITLE OF INVENTION: METHODS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Heller Ehrman White & McAuliffe  
; STREET: 4250 Executive Square, Suite 700  
; CITY: La Jolla  
; STATE: California  
; COUNTRY: US  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/984, 709A  
; FILING DATE: 02-DEC-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seidman, Stephanie L.  
; REGISTRATION NUMBER: 33,779  
; REFERENCE/DOCKET NUMBER: 24735-9815 (formerly 6362-9815)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 450-8400  
; TELEFAX: (619) 587-5360  
; INFORMATION FOR SEQ ID NO: 23:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7791 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 237..7037  
; OTHER INFORMATION: /standard\_name= "Alpha-1A-2"  
US-08-984-709A-23

Alignment Scores:  
Pred. No.: 0.0401 Length: 7791  
Score: 161.00 Matches: 165  
Percent Similarity: 29.60% Conservative: 73  
Best Local Similarity: 20.52% Mismatches: 234  
Query Match: 3.26% Indels: 334  
DB: 4 Gaps: 41

US-09-720-086-7 (1-912) x US-08-984-709A-23 (1-7791)  
Qy 7 SerGlyProGlyAspThr-----SerSerAlaAlaGluArgGluGlu 21  
Db 2636 AATGGACCCGACGAGCGCTGGAAGGCTGCTACACGCGCGACCTCGCGCCACATGAA 2695  
Qy 22 Asp-----ArgLysAspGlyGluGluGluGluGluGluGluGlu 37

Db 2696 GAGCAGCTTGGACCGCGCGCTGTGTGTGGAGCCGACAGAGAACCGCAA---CAACAACAC 2752  
Qy 38 ArgGlnGluProSerThrThrAlaArgLys-----ValGlyArgProGlyArgLysArg 55  
Db 2753 CAACAAGAGCG 2812  
Qy 56 LysHisProProValGluSerGlyAspThrProLysAspProAlaValLleSerLysSer 75  
Db 2813 GGACTTCTCTAG---GAAACAGCGCGCGCTACACGATCG-----GGCCCGGGA 2857  
Qy 76 ProSerMetAlaGlnAspSerGlyAlaSerGluLeuLeu-----ProAsnGlyAsp 92  
Db 2858 CCCACAGCGCTCGCGCGCGCTGACGACGACGAGCGCTGGCGCGGAAGCCAGGAGCGCA 2917  
Qy 93 LeuGluLysArgSerGluProGlnProGluGluGluSerProAlaGlyGlyGlnLysGly 112  
Db 2918 GTGAGCGCGGAGGACCTTACCGCGCGAGTCGACCA-----CGCCCGGA 2968  
Qy 113 GlyAlaProAlaGluGlyGluGlyAlaAlaGluThrLeuProGluAlaSerArgAlaVal 132  
Db 2969 GGGCAGCCT-----GGAGCAACCCGGGTCTCGGAGGCGGAGCGCGAGCG--- 3013  
Qy 133 GluAsnGlyCysCysThrProLysGluGlyArgGlyAlaProAlaGlu----- 148  
Db 3014 -----AGGCAAGCGCGGACCCCGAGGACGTCGACCG 3052  
Qy 149 AlaGlyLysGluGlnLysGluThrAsn----- 157  
Db 3053 GCAGGGCGCAGCAGGAGAGCGCGCGCTCCCGCAGCGCGCGCGCGCGGAGCA 3112  
Qy 158 IleGluSerMetLysMetGluGlySerArgGlyArgLeuArgGlyGlyLeuGlyTrpGlu 177  
Db 3113 TCGAGCTCATCGCGCACCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3172  
Qy 178 SerSerLeuArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTyrTyr 197  
Db 3173 GAGGCGCGCGCACCGCGA-----GGGACCGCGCGCGCGCGCGCGCGCGCGCGCG 3205  
Qy 198 IleSerLysArgLysArgAspGluTrpLeuAlaArgTrpLysArgGluAlaGluLysLys 217  
Db 3206 CCGGGCG 3259  
Qy 218 AlaLysVal-----IleAlaGlyMetAsnAlaValGluGluAsnGlnGly 232  
Db 3260 GCACCGCATGCGCTCCAGCCAGCCAGCTACGAGGGGCGCGCGCGCGCGCGCGCGCGCG 3316  
Qy 233 ProGlyGluSerGlnLysValGluGluAlaSerProProAlaValGlnGlnProThrAsp 252  
Db 3317 -----GCGAGGCGATCGGAGGAGAGAAAGAGAACAGGGCTCCGGGTCCC----- 3361  
Qy 253 ProAlaSerProThrValAlaThrThrProGluProValGlySerAspAlaGlyAspLys 272  
Db 3362 -----TGTGTGCGGCGCGCAACCTGTC-----AACAC 3388  
Qy 273 AsnAlaThrLysAlaGlyAspAspGluProGluTyrGluAspGlyArgGlyPheGlyIle 292  
Db 3389 CCGGCCAATCCAGCAGCGACCTGGCGCGCGCAAGACCCCGCGCGCGCGCGCGCGCGCGCG 3448  
Qy 293 GlyGluLeuValTrpGlyLysLeuArgGlyPheSerTrp-----TrpPro 307  
Db 3449 CATGAAGAACAACAGCTCGCCACCGCGGAGTCGCGCGCTCCCGACCGCAGCTTGGCA 3508  
Qy 308 GlyArgIleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGluGlyThrArgTrp 327  
Db 3509 -----CGCCGCGCTCG 3535  
Qy 328 ValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLysLeuMetProLeu 347  
Db 3536 -----GATGGAAA----- 3544  
Qy 348 SerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnProMetTyrArgLys 367  
Db 3545 -----CAGCAGCG 3562

QY	368	AlaIleTyrGluValLeuGlnValAlaIaSerSerArgAlaGlyIleuLeuPheProValCys	387
Db	3563	-----CATGCTGGCCAT-----CCCTGC	3580
QY	388	HisAspSerAspGlnSerAspThrAlaValAlaValGluValGlnAsnIleProMetIle	407
Db	3581	CATGGCCACCAACCCCGAGAACGCCGCCGACGCCGAGCCCA-----	3625
QY	408	GluThrAlaLeuGlyGlyPheGlnProSerGlyProIleuGluProProGluGln	427
Db	3626	-----CAACCCGGGGAAACCATTCCAATCCCGGCCCCCCCAAGAC	3664
QY	428	GluIleAsn---ProTyrIleS-----GluValTyrThrAspMetTrp-----	440
Db	3665	CCCCGAGAAATAGCCTTATCGTCACCAACCCGAGCGGACCCAGACCAATTCAGTTAAGAC	3724
QY	441	-----ValGluProGluAlaAlaAlaTyrAlaPro-----ProProAlaIleS	455
Db	3725	TGCCAGAAACCCGACACACACCAAGTAGACATCCCCGAGCTGGCCACCCCTCA	3784
QY	456	LysProAlaGlySerThrAlaGluIleProIleValIleValGluIleLeuAspGluThr	475
Db	3785	CCACACCGTCGTACAGTGAACAAACGCAACCCAGACCCACTGCGCAAAAAAGAGA	3844
QY	476	ArgGluArgLeuValTyrGluValArgGluIleCysArgAsnIleGluAspIleCysIle	495
Db	3845	AGAGAAAGAA-----	3853
QY	496	SerCysGlySerLeuAsnValThrLeuGlnHisProLeuPheValGlyIleMetCysGln	515
Db	3854	-----GGAGGA-----	3859
QY	516	AsnCysIleAsnCysPheLeuGluCysAlaTyrGlnTyrAspAspArgIleTyrGlnSer	535
Db	3859	-----	3859
QY	536	TyrCysThrIleCysCysGlyGlyArgGluValLeuMetCysGlyAsnAsnAsnCysCys	555
Db	3860	-----GGAGGAAGAGA-----	3871
QY	556	ArgCysPheCysValGluCysValaAspLeuLeuValGlyProGluAlaAlaGlnAlaAla	575
Db	3871	-----	3871
QY	576	IleIleValAspProTyrAsnCysTyrMetCysGlyHisIleGlyThrTyrGlyLeuLeu	595
Db	3872	-----CCGTGGGA-----	3880
QY	596	ArgArgArgGluAspTrpProSerArgLeuGlnMetPhePheAlaAsnAsnHisAspGln	615
Db	3881	---AGACGGCCCTAAGCAATGCCCTCCCTATAGCTCCATGTTATCCGTCCACGACCA	3937
QY	616	GluPheAspProProIleValTyrProProValProAla-----	628
Db	3938	CCCCCTTGGCGCTGTGCCATTACCTCTGAACCTGGCTACTTTGAGATGTGATCCT	3997
QY	629	-----GluIleArgIleProIleArg-----	635
Db	3998	CATGTCATTGCCATGACGACGATCGCCCTGGCCGCGAGACCCCTGCGACCCCAAGC	4057
QY	636	-----ValLeuSerLeuPheAspGlyIleAlaThrGlyLeuLeuValLeuIle	651
Db	4058	ACCTGGGAACAGCTGCTCGCATCTTGACTACGTTTTCACAGCGCTTACCTTT--	4115
QY	651	AspLeuGlyIleGlnValaAspArgTyrIleAlaSerGluValCysGluAspSerIleThr	671
Db	4116	-----GAGATGTGATCAAGATGATTGA	4138
QY	671	IleValIleMetValaArgHisGlnGlyIleIleMetTyrValGlyAspValArgSerValThr	691
Db	4139	CCTGGGCTGTCTCTGATCAAGGT-----GCTTACTTCCGTAACTCTGGAATATTCT	4192

```

QY 691 RGLNtYHSHSllEGInGtUTPQlYPrPhEAsPLeuValllEGlYglYsErPrOCySAs 711
Db 4193 -CGACTTCAATAGTGCTGACGTGGGCCCT-----GTAAGCTTTGGCCT 4233
QY 711 nAsPLeuSerllE-----ValAnPrOAlAaG---LySGl 722
Db 4234 -TCACtTGCAATAGCAAAAGAAAGACATCAACACGATTAAATCCCTCGAGTCTCCGG 4292
QY 722 YLeuTYrGlu 725
Db 4293 GTGCTAGCAC 4302

RESULT 14
US-08-450-272-23
; Sequence 23, Application US/08450272
; Patent No. 6387696
; GENERAL INFORMATION:
; APPLICANT: Harpold, Michael
; APPLICANT: Ellis, Steven
; APPLICANT: Williams, Mark
; APPLICANT: McCue, Ann
; APPLICANT: Gillespie, Alison
; APPLICANT: Feldman, Daniel
; APPLICANT: Brenner, Robert
; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: California
; COUNTRY: US
; ZIP: 92101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,272
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/404,950
; FILING DATE: 13-MAR-1995
; APPLICATION NUMBER: 08/336,257
; FILING DATE: 7-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/314,083
; FILING DATE: 28-SEPT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/311,363
; FILING DATE: 23-SEPT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290,012
; FILING DATE: 11-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/223,305
; FILING DATE: 4-APR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/193,078
; FILING DATE: 07-FEB-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/149,097
; FILING DATE: 5-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/105,536
; FILING DATE: 11-AUG-1993

```

```

, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: 07/914,231
, FILING DATE: 13-JULY-1992
, CLASSIFICATION: 435
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: 07/868,354
, FILING DATE: 10-APR-1992
, CLASSIFICATION: 435
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: PCT/US92/00
, FILING DATE: 14-AUG-1992
, CLASSIFICATION: 435
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: 07/745,206
, FILING DATE: 15-AUG-1991
, CLASSIFICATION: 435
, ATTORNEY/AGENT INFORMATION:
, NAME: Seidman, Stephanie L.
, REGISTRATION NUMBER: 33,779
, REFERENCE/DOCKET NUMBER: 6362-5-1
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: (619) 238-0999
, TELEFAX: (619) 238-0062
, INFORMATION FOR SEQ ID NO: 23:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 7791 base pairs
, TYPE: nucleic acid
, STRANDEDNESS: double
, TOPOLOGY: linear
, MOLECULE TYPE: DNA (genomic)
, FEATURE:
, NAME/KEY: CDS
, LOCATION: 237..7037
, OTHER INFORMATION: /standard_name=
US-08-450-272-23

```

OTHER INFO  
US-08-450-272-23

Alignment Scores:

Pred. No.:	0.0401	Length:	7791
Score:	161.00	Matches:	165
Percent Similarity:	19.60%	Conservative:	73
Best Local Similarity:	20.52%	Mismatches:	234
Query Match:	3.26%	Indels:	334
DB:	4	Gaps:	41

US-09-720-086-7 (1-912) x US-08-450-272-23 (1-7791)

Qy	7	SerGlyProGlyAspThr	-----SerSerSerAlaAlaGluArgGluGlu	21
Db	2636	AATGGACCGGACGAGCGCTGGAAGGCTGCCTACACGUGCGGACCTCGCGCCAGACATGAA	2695	
Qy	22	Asp	-----ArgLysAspGlyGluGluGluGluGluProArgGlyLysGluGlu	37
Db	2696	GACGCACCTTGACCGCGCGCTGGTGGACCCGCGAGGAGAACCGCAA	2752	
Qy	38	ArgGlnGluProSerThrThraAlaArgLys	-----ValGlyArgProGlyArgLysArg	55
Db	2753	CAACAAGACCGGGCGCGCCGAGCCACCCTGGACACGAGCGCTCGGCCACGACGCGGCCGA	2812	
Qy	56	LysHisProProValGluSerGlyAspThrProLysAspProAlaValIleSerLysSer	75	
Db	2813	GGACTTCCTCAG--GAACAACGCGCGCTACCCAGCATCG	2857	
Qy	76	ProSerMetAlaGlnAspSerGlyAlaSerGluLeuLeu	-----ProAsnGlyAsp	92
Db	2858	CCCCAGCGGCTCGCGGGCTGGACGCGACGAGGCGCTGGCGGGNAGCCAGAGGCCGA	2917	
Qy	93	LeuGlnLysArgSerGluProGlnProGluGluGlySerProAlaGlyGlyGlnLysGly	112	
Db	2918	GCTGAGCCGGAGGAGACCTACGCGCGGATCGGACACCA	2968	
Qy	113	GlyAlaProAlaGluGlyGluGlyAlaAlaGluThrLeuProGluAlaSerArgAlaVal	132	
Db	2969	GGGCAGCGCT-----GGAGCAACCGGGTTCTGGGAGGCGGAGCGCGACG	3013	



QY 76 ProSerMetAlaGlnAspSerGlyAlaSerGluLeuLeu-----ProAsnGlyAsp 92  
Db 2858 CCCAGCGGCTCGCGGGCTGGACCGACCGAGCGCTCGCGGGGAAGCCAGGAGGCGGA 2917  
QY 93 LeuGluLysArgSerGluProGlnProGluGluGlySerProAlaGlyGlyGluLysGly 112  
Db 2918 GCTGAGCCGGAGGAGCCCTACGGCCGCGAGTCGACCA-----CGCCCGGA 2968  
QY 113 GlyAlaProAlaGluGlyGluGlyAlaAlaGluThrLeuProGluAlaSerArgAlaVal 132  
Db 2869 GGGCAGCCT-----GGAGCAACCCGGGTCTGGAGGCGCAGGCGGAGCG--- 3013  
QY 133 GluAsnGlyCysCysThrProLysGluGlyArgGlyAlaProAlaGlu----- 148  
Db 3014 -----AGGCAAGCCGGGACCCCGACCGAGGCACGTGCACCG 3052  
QY 149 AlaGlyLysGluGlnLysGluThrAsn----- 157  
Db 3053 GCAGGGGGCAGCAGGAGCGCGAGCGGTCCCGCGCACGGCGCGGACGGAGCA 3112  
QY 158 IleGluSerMetLysMetGluGlySerArgGlyArgLeuArgGlyGlyLeuGlyTrpGlu 177  
Db 3113 TCGAGCTCATCGCGCGCAGCGAGGCGCGGGAGGAGGTCCGGAGGACAAAGCGGAGCG 3172  
QY 178 SerSerLeuArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTyrTyr 197  
Db 3173 GAGGCGCGGCACCGCA-----GGCGCCCGCGCGCGC 3205  
QY 198 IleSerLysArgLysArgAspGluTrpLeuAlaArgTrpLysArgGluAlaGluLysLys 217  
Db 3206 CCGGGCGCGGAGGCGAGGCGCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3259  
QY 218 AlaLysVal-----IleAlaGlyMetAsnAlaValGluGluAsnGlnGly 232  
Db 3260 GCACCGCATGCGCTCCAGCCAGTACGAGGGGCGCGCG-----GAGGGAGGACAAGGA 3316  
QY 233 ProGlyGluSerGlnLysValGluGluAlaSerProProAlaValGlnGlnProThrAsp 252  
Db 3317 -----GCGAGGCATCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3361  
QY 253 ProAlaSerProThrValAlaThrThrProGluProValGlySerAspAlaGlyAspLys 272  
Db 3362 -----TGTGTCGGGCGCCCACTGTC-----AACAC 3388  
QY 273 AsnAlaThrLysAlaGlyAspAspGluProGluTyrGluAspGlyArgGlyPheGlyIle 292  
Db 3389 CCGGCCAATCCAGCAGGACCTGGCGCGCCCAAGACCCACCCCTGGCAGAGGATATGACAA 3448  
QY 293 GlyGluLeuValTrpGlyLysLeuArgGlyPheSerTrp-----TrpPro 307  
Db 3449 CATGAAGAACAAAGCTGGCCACCGCGGAGTCGGCGCTCCCGCACGGCAGCCTTGGCCA 3508  
QY 308 GlyArgIleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGluGlyThrArgTrp 327  
Db 3509 -----CGCCGGCTGCCCGCAGAGCCAGCCAA--- 3535  
QY 328 ValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLysLeuMetProLeu 347  
Db 3536 -----GATGGAAA----- 3544  
QY 348 SerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnProMetTyrArgLys 367  
Db 3545 -----CAGCACCGACCCCGGCC----- 3562  
QY 368 AlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeuPheProValCys 387  
Db 3563 -----CATCTGGCCAT-----CCCTGC 3580  
QY 388 HisAspSerAspGluSerAspThrAlaLysAlaValGluValGlnAsnLysProMetIle 407  
Db 3581 CATGCCCAACCAACCG 3625

QY 408 GluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGluProGluGlu 427  
Db 3626 -----CAACCCGGGAACCCATCCATCCCGCCCCCAAGAC 3664  
QY 428 GluLysAsn---ProTyrLys-----GluValTyrThrAspMetTrp----- 440  
Db 3665 CCGCGAGAATAGCTTATCGTCACCAACCCCGGCGGACCCAGACCAATTACGTAAAGAC 3724  
QY 441 -----ValGluProGluAlaAlaAlaTyrAlaPro-----ProProAlaLys 455  
Db 3725 TGCAGGAAACCCGACACACACAGTGGACATCCCGACGCTGCCACCCCTCTCAA 3784  
QY 456 LysProArgLysSerThrAlaGluLysProLysValLysGluLysIleAspGluArgThr 475  
Db 3785 CCACACCGTGTACAAAGTGAACAAACGCAACCCAGACCCACTGCCAAAAAAGAGGA 3844  
QY 476 ArgGluArgLeuValTyrGluValArgGlnLysCysArgAsnIleGluAspIleCysIle 495  
Db 3845 AGAGAAGAA----- 3853  
QY 496 SerCysGlySerLeuAsnValThrLeuGluHisProLeuPheValGlyGlyMetCysGln 515  
Db 3854 -----GGAGGA----- 3859  
QY 516 AsnCysLysAsnCysPheLeuGluCysAlaTyrGlnTyrAspAspAspGlyTyrGlnSer 535  
Db 3859 ----- 3859  
QY 536 TyrCysThrIleCysCysGlyGlyArgGluValLeuMetCysGlyAsnAsnAsnCysCys 555  
Db 3860 -----GGAGGAAGACGA----- 3871  
QY 556 ArgCysPheCysValGluCysValAspLeuValGlyProGlyAlaAlaGlnAlaIle 575  
Db 3871 ----- 3871  
QY 576 IleLysGluAspProTyrAsnCysTyrMetCysGlyHisLysGlyThrTyrGlyLeuLeu 595  
Db 3872 -----CCGTGGGA----- 3880  
QY 596 ArgArgArgGluAspTrpProSerArgLeuGlnMetPhePheAlaAsnAsnHisAspGln 615  
Db 3881 ---AGAGCGCCCTAAGCCAATGCTCCCTATAGTCCATGTTCTCCTGTCACGACCA 3937  
QY 616 GluPheAspProProLysValTyrProProValProAla----- 628  
Db 3938 CCCCCTTCGCGCGCTGCGCATTCATCATCTGAACTCGGCTACTTTGAGATGTGATCCT 3997  
QY 629 -----GluLysArgLysProIleArg----- 635  
Db 3998 CATGTCATTCGATGAGCAGCATCGCCCTGCGCGGAGGACCCCTGTGACGCCCAACGC 4057  
QY 636 -----ValLeuSerLeuPheAspGlyIleAlaThrGlyLeuValLeuLys 651  
Db 4058 ACCTCGGAACAACGCTGCTGCGATCTTTGACTAGCTTTTACAGCGCTCTTCCACCTTT-- 4115  
QY 651 AspLeuGlyIleGlnValAspArgTyrIleAlaSerGluValCysGluAspSerIleTh 671  
Db 4116 -----GAGATGGTGATCAAGATGATTGA 4138  
QY 671 rValGlyMetValArgHisGlnGlyLysIleMetTyrValGlyAspValArgSerValTh 691  
Db 4139 CTTGGGGCTGCTCTGATCAGGT-----GCCTACTTCGCTGACCTCTGGAATATCT 4192  
QY 691 rGlnLysHisIleGlnGluTrpGlyProPheAspLeuValIleGlyGlySerProCysAs 711  
Db 4193 -CGACTTCATAGTGTGTCAGTGGGCGCT-----GGTAGCTTTGGCT 4233  
QY 711 nAspLeuSerIle-----ValAsnProAlaIleArg---LysGlu 722  
Db 4234 -TCACTGGCAATAGCAAGGAAAGACATCAACAGATTAAATCCCTCCGAGTCTCTCCGG 4292  
QY 722 yLeuTyrGlu 725

Mon Nov 25 08:18:36 2002

us-09-720-086-7.rn1

Page 23

Db 4293 GTGCTACGAC 4302

Search completed: November 25, 2002, 02:59:49  
Job time : 272.557 secs



**THIS PAGE BLANK (USPTO)**

AF067972 3005 bp mRNA linear PRI 12-FEB-2001  
 LOCUS AF067972 3005 bp mRNA linear PRI 12-FEB-2001  
 DEFINITION Homo sapiens DNA cytosine methyltransferase 3 alpha (DNMT3A) mRNA,  
 complete cds.  
 ACCESSION AF067972  
 VERSION AF067972.2 GI:12746531  
 KEYWORDS  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 Xie,S., Wang,Z., Okano,M., Nogami,M., Li,Y., He,W.W., Okumura,K.  
 and Li,E.  
 Cloning, expression and chromosome locations of the human DNMT3  
 gene family  
 Gene 236 (1), 87-95 (1999)  
 PUBMED 10433969  
 REFERENCE 2 (bases 1 to 3005)  
 Xie,S., Okano,M. and Li,E.  
 Direct Submission  
 Submitted (25-MAY-1998) CVRC, Mass. General Hospital, 149 13th  
 Street, Charlestown, MA 02129, USA  
 3 (bases 1 to 3005)  
 Hata,K., Shirohzu,H., Sasaki,H. and En,L.  
 Direct Submission  
 Submitted (12-FEB-2001) CVRC, Mass. General Hospital, 149 13th  
 Street, Charlestown, MA 02129, USA  
 REMARK  
 COMMENT Sequence update by submitter  
 On Feb 12, 2001 this sequence version replaced GI:4927369.  
 FEATURES  
 source  
 1..3005  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2p23; between D2S171 and D2S174; 48-50 cM"  
 1..3005  
 /gene="DNMT3A"  
 230..2968  
 /gene="DNMT3A"  
 /function="de novo DNA cytosine methyltransferase"  
 /codon\_start=1  
 /product="DNA cytosine methyltransferase 3 alpha"  
 /protein\_id="AA033084.2"  
 /db\_xref="GI:12746532"  
 /translation="MPAPSSGPGDTSSAAREEDRKDGEDEPRKGEKEOEPEPTT  
 ARVGRGRKRKHPRVSGDTPKDPKPAVSKSPMAODSGSELINPGLKRSSEPE  
 EGSPPAGGOKGAPAGEGAETLPEASRAVENGCCTPKERGAFAEAKOEKTNIES  
 MKKEGSRGLRGGLWESSLRPMPRLTFOAGDPYISRKDEWLARKWEAKKA  
 KVLGNNAVENOGESOKVEASPAVQPTDPAVATTPPEVPSDAGDKATK  
 AGDPEYEDRGFGIGELVMGKLRFSGMWRGRIYMMWGRSRAAEGTRWMPDGE  
 KESVCEKMLPSFSCSAFHQATYKQPMRKATYEVLOVASRAGLFPVCHSDS  
 SDTAKAVEONKPMLEWALGFGFOPSGKLEPPEERKNPKYEVITDMVEPAAATP  
 PPPAKPKRSKTAEKPKKEIIDERTRELVYEVROKRNIEDICISCSGLNTLEHL  
 FVGMCQNKCEFLCAGYDDYDGYOCTICCGREVMGNNGCCFCVCEVDL  
 VGGGAQAAIKEDPNCVMGKGTGLIRREDMPSRLQMFANNDHOEPKPKYP  
 PVPBAEKRTIVLSLFDGATGLVLKDLGIQVDRIYASEVCESDITGVNRHGGKIM  
 YVGDVRSVTRQHIQEWGPFDLVIGGSPCNDSIVNPARGLYEGTGLFEFRLIAD  
 ARPKEDDRPFVLFENVVAMGVSDKRDISRLFESNPVIMDAKEVSAHARAYFWNL  
 PKNRPPLASTVNDKLEOCLEHGRIAKFSKRYRTITRSNIDIKOGKHPPVPMKE  
 DILNCEMERFPGFVHYTDVSNMSRLARKRLGLGRSWSVPIHLLFAPLKEYFACV"

BASE COUNT 674 a 846 c 978 g 507 t

Alignment Scores:  
 Pred. No.: 1.67e-188  
 Score: 4935.00  
 Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 100.00%  
 DB: 9 Gaps: 0

Length: 3005  
 Matches: 912  
 Conservative: 0  
 Mismatches: 0  
 Indels: 0  
 Gaps: 0

US-09-720-086-7 (1-912) x AF067972 (1-3005)  
 QY 1 MetProAlaMetProSerSerGlyProGlyAspThrSerSerAlaAlaGluArgGlu 20  
 Db 230 ATGCCCGCCATTGCCCTCCAGCGGCCCCGGGGACACACAGCTCTCTCTCCGAGCGGAG 289  
 QY 21 GluAspArgLysAspGlyGluGluGluGluGluProArgGlyLysGluGluArgGluGlu 40  
 Db 290 GAGGACGGAAG 349  
 QY 41 ProSerThrThrAlaArgLysValGlyArgProGlyArgLysArgLysHisProProVal 60  
 Db 350 CCAGACACACAGCGACGAG 409  
 QY 61 GluSerGlyAspThrProLysAspProAlaValIleSerLysSerProSerMetAlaGln 80  
 Db 410 GAAAGCGGTACACAGCCAAAG 469  
 QY 81 AspSerGlyAlaSerGlyLysLysLysLysLysLysLysLysLysLysLysLysLysLys 100  
 Db 470 GACTCAGCGCCCTCAGAGCTATTACCAATGGGAGCTTGAGAGAGAGAGAGAGAGAGAGAG 529  
 QY 101 ProGluGluGluSerProAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 120  
 Db 530 CCAG 589  
 QY 121 AlaAlaGluThrLeuProGluAlaSerArgAlaValGluSerGlyCysCysThrProLys 140  
 Db 590 GAGCTGAG 649  
 QY 141 GluGlyArgGlyAlaProAlaGluAlaGlyLysGluGluGluGluGluGluGluGluGlu 160  
 Db 650 GAGGCGGAG 709  
 QY 161 MetLysMetGluGlySerArgGlyArgLysArgGlyGlyLysGlyLysGlyLysGlyLys 180  
 Db 710 ATGAATGAG 769  
 QY 181 ArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProGlyArgLysSerLys 200  
 Db 770 GCTCAG 829  
 QY 201 ArgLysArgAspGluThrPheLysAlaArgGlyLysArgGlyLysAlaGluLysLysAlaLysVal 220  
 Db 830 CGAAG 889  
 QY 221 IleAlaGluMetSerAlaValGluGluGluGluGluGluGluGluGluGluGluGluGlu 240  
 Db 890 ATTGAG 949  
 QY 241 GluAlaSerProProAlaValGluGluGluGluGluGluGluGluGluGluGluGluGlu 260  
 Db 950 GAGGCGAG 1009  
 QY 261 ThrProGluProValGlySerAspAlaGlyAspLysAspAlaThrLysAlaGlyAspAsp 280  
 Db 1010 AGCGCTGAG 1069  
 QY 281 GluProGluThrArgLysAspGlyArgGlyPheGlyIleGlyGluLysValTTPGlyLysLeu 300  
 Db 1070 GAGCGAG 1129  
 QY 301 ArgGlyPheSerThrProGluArgLysIleValSerThrProMetThrGlyArgSerArg 320  
 Db 1130 CGGGGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1189  
 QY 321 AlaAlaGluGlyThrArgTrpValMetTrpPheGlyAspGlyLysPheSerValAlaCys 340  
 Db 1190 GACAGTGAAG 1249  
 QY 341 ValGluLysLeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTrpAsn 360  
 Db 1250 GTTGAAG 1309



QY 361 LysGlnProMetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAla 380  
 DB 1310 AAGCAGCCCACTGATACCGCAAGCAGCATCTGACAGGTCTGACAGGTGGCCGACGCGCGC 1369  
 QY 381 GlyLysLeuPheProValCysHisAspSerAspGluSerAspThrAlaLysAlaValGlu 400  
 DB 1370 GCGAAGCTGTTCCCGGTGGCCAGCAGCATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1429  
 QY 401 ValGlnAsnLysProMetIleGluTrpAlaLeuGlyGlyPheGlnProSerGlyProLys 420  
 DB 1430 GTGCAAGAACAGCCCATGATTTGATGAGTGGCCCTGGGGGGCTTCCAGCCTTCTGGCCCTAAG 1489  
 QY 421 GlyLeuGluProProGluGluGluLysAsnProTyrLysGluValTyrThrAspMetTrp 440  
 DB 1490 GGGCTGAGAGCCCAAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAACAGAAC 1549  
 QY 441 ValGluProGluAlaAlaAlaTyrAlaProProProProAlaLysLysProArgLysSer 460  
 DB 1550 GTGGAAGCTGAGGAGCTGCTCCCTGACCACTCCACAGCCCAAAAAGCCCGGAAAGAGC 1609  
 QY 461 ValGluLysProLysValLysGluIleLeuAspGluArgThrArgGluVal 480  
 DB 1610 AAGCGAGAACCCCAAGCTCAAGAGATTTATGATGAGCGCACAGAGAGCGCTGTG 1669  
 QY 481 TyrGluValArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeu 500  
 DB 1670 TACGAGGTGGCGGAGAGAGTGGCGAACATGAGACATGATCTGCTGTGGAGAGCTC 1729  
 QY 501 AsnValThrLeuGlnHisProLeuPheValGlyLysMetCysGlnAsnCysLysAsnCys 520  
 DB 1730 AATGTTACCTCGGAACACCCCTCTCTGTGGAGAGATGTGCCAAACTGCAAGAACTGC 1789  
 QY 521 PheLeuGluCysAlaTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrIleCys 540  
 DB 1790 TTCTGAGTGTGCTGACAGTACGAGCAGCAGCAGCAGCAGTCTGACATCTGC 1849  
 QY 541 CysGlyLysArgGluValLeuMetCysGlyValAsnAsnAsnCysCysArgCysPheCysVal 560  
 DB 1850 TGTGGGGGCGCTGAGGTGTCTCATGTGCGGAAACAACTGCTGAGGTCTTTGCTG 1909  
 QY 561 GluCysValAspLeuLeuValGlyProGlyAlaAlaGlnAlaIleLysGluAspPro 580  
 DB 1910 GAGTGTGTGACCTCTTGTGGGGCGCGGCGCTGCCAGGACGACATTAAGAAAGACCCC 1969  
 QY 581 TrpAsnCysTyrMetCysGlyHisLysGlyThrTyrGlyLeuArgArgArgGluAsp 600  
 DB 1970 TGGAACTGTCTACATGTGCGGCGCAAGGCTTACCTGCGGCTGTGCGGCGGAGAGAC 2029  
 QY 601 TrpProSerArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProPro 620  
 DB 2030 GGGCTGCGCGCTCCAGAGTCTTCTGTAATACCAAGACAGAAATTTGACCCCTCCA 2089  
 QY 621 LysValTyrProProValProAlaGluLysArgLysProIleArgValLeuSerLeuPhe 640  
 DB 2090 AAGGTTAACCCACCTGTCTCCAGCTGAGAGAGAGAACCCCATCCGGGTCTGTCTCTTT 2149  
 QY 641 AspGlyIleAlaThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyr 660  
 DB 2150 GATGGAAATCGCTACAGAGCTCTCTGCTGAAAGACTTGGGCAATTCAGGTGACCGCTAC 2209  
 QY 661 IleLeuAspGluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLys 680  
 DB 2210 ATTGCTCTGAGAGTGTGTGAGATCTTCATCAGGTGGCATGTGTGGGACCAAGGAGAG 2269  
 QY 681 IleMetTyrValGlyAspValArgSerValThrGlnLysHisIleGlnGluTrpGlyPro 700  
 DB 2270 ATCATGTACGTGGGAGAGCTCGGAGCTCACACAGAGCATATCCAGAGTGGGGCCA 2329  
 QY 701 PheAspLeuValIleGlyLysSerProCysAsnAspLeuSerIleValAsnProAlaArg 720  
 DB 2330 TTGATCTGTGATTTGGGGGAGTCCCTGCAATGACCTCTCCATCTGTCAACCTGCTGCGC 2389

QY 721 LysGlyLeuTyrGluGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAsp 740  
 DB 2390 AAGGCGCTTACAGAGGAGCATGGCCGGCTCTTCTTGAGTTCTTACCGCTCTGCTGATAT 2449  
 QY 741 AlaArgProLysGluGlyAspAspArgProPhePheTyrPhePheGlnAsnValAla 760  
 DB 2450 GCGCGGCGCCAGGAGGAGATGATCGCCCTTCTTGTGCTCTTGAAGAATGTGGTGGC 2509  
 QY 761 MetGlyValSerAspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIle 780  
 DB 2510 ATGGCGCTTATGTGACMAAGGAGCATCTCGCATTTCTGAGTCCAAACCTGTGATGATT 2569  
 QY 781 AspAlaLysGluValSerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGly 800  
 DB 2570 GATGCCAAGAGATGTGACGTGACACAGAGGCGCGCTACTTCTGGGTAACTCTCCCGGT 2639  
 QY 801 MetAsnArgProLeuAlaSerThrValAsnAspLysLeuGluLeuGlnGluCysLeuGlu 820  
 DB 2630 ATGAACAGGCGGTGGCATCTCCATGATGATGATGATGATGATGATGATGATGATGATGATG 2689  
 QY 821 HisGlyArgIleAlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIle 840  
 DB 2690 CATGGCAGATAGCCCAAGTTCAGCAAGTGAAGACCATTAATCAAGATCAAGATCAAGATCA 2749  
 QY 841 LysGlnGlyLysAspGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTrp 860  
 DB 2750 AAGCAGGCGCAAGACCAAGCATTTCTGCTTCTTCAATGAAATGAGAAAGAGACATCTTATGG 2809  
 QY 861 CysThrGluMetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMet 880  
 DB 2810 TGCACGTGAATGGAAGGATATTGTTTCCAGTCCACTAATCTGACGCTCTCCAACTG 2869  
 QY 881 SerArgLeuAlaArgGluArgLeuLeuGlyArgSerTrpSerValProValIleArgHis 900  
 DB 2870 AGCCGCTGGGAGGAGAGACTGTGGCCGGTGTATGAGAGCTGCTCCAGTATCCGCCAC 2929  
 QY 901 LeuPheAlaProLeuLysGluLysPheAlaCysVal 912  
 DB 2930 CTCTTCCTCCGCTGAAGAGATTTTTCGCTGTG 2965

RESULT 2  
 AF331856  
 ----- 4250 bp mRNA 1 bp 5' 3'

**THIS PAGE BLANK (USPTO)**

13	125.5	2.7	4411529	4	US-09-103-840A-1	Sequence 1, Appl
14	125	2.7	4657	3	US-09-254-325-1	Sequence 1, Appl
15	125	2.7	5643	3	US-09-079-415-5	Sequence 5, Appl
16	125	2.7	5643	3	US-08-750-458A-1	Sequence 1, Appl
17	124	2.7	4181	4	US-09-393-569-1	Sequence 1, Appl
18	123.5	2.7	28804	2	US-08-592-874-1	Sequence 1, Appl
19	123.5	2.7	28804	2	US-09-096-942-2	Sequence 2, Appl
20	123.5	2.7	28804	2	US-09-096-942-2	Sequence 2, Appl
21	123.5	2.7	403765	4	US-09-103-840A-2	Sequence 2, Appl
22	123	2.7	3044	4	US-09-152-060-36	Sequence 36, Appl
23	123	2.7	36519	3	US-08-923-137-2	Sequence 1, Appl
24	122	2.7	8931	3	US-09-051-019-1	Sequence 1, Appl
25	121.5	2.6	5965	4	US-09-362-336A-1	Sequence 1, Appl
26	121.5	2.6	6028	4	US-09-362-336A-3	Sequence 3, Appl
27	121.5	2.6	20235	1	US-07-642-734C-3	Sequence 3, Appl
28	121.5	2.6	20235	1	US-08-439-009A-3	Sequence 3, Appl
29	120.5	2.6	1170	3	US-09-307-621-1	Sequence 1, Appl
30	119	2.6	11471	4	US-09-504-358-16	Sequence 16, Appl
31	118.5	2.6	11471	4	US-09-504-314-16	Sequence 16, Appl
32	118.5	2.6	1865	4	US-09-370-253-5	Sequence 5, Appl
33	118.5	2.6	6463	2	US-08-962-284-3	Sequence 3, Appl
34	118	2.6	1970	1	US-08-028-463-14	Sequence 14, Appl
35	118	2.6	1970	1	US-08-461-836-14	Sequence 14, Appl
36	118	2.6	43280	2	US-08-804-227C-1	Sequence 1, Appl
37	117.5	2.6	11604	4	US-09-385-028-13	Sequence 13, Appl
38	117.5	2.6	15079	4	US-09-385-028-1	Sequence 1, Appl
39	117	2.5	2905	4	US-08-942-686-1	Sequence 1, Appl
40	117	2.5	4265	4	US-09-061-709-1	Sequence 1, Appl
41	116.5	2.5	43280	2	US-08-804-227C-1	Sequence 1, Appl
42	116	2.5	49377	1	US-08-764-233A-1	Sequence 1, Appl
43	115.5	2.5	1800	1	US-08-139-937-11	Sequence 11, Appl
44	115.5	2.5	1800	5	PCT-US93-11310-11	Sequence 11, Appl
45	115	2.5	50937	4	US-09-428-517-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-09-276-531-47  
Sequence 47, Application US/09276531  
Patent No. 6183968  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Yee, Henry  
APPLICANT: Reddy, Roopa  
APPLICANT: Guejler, Karl J.  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
NUMBER OF SEQUENCES: 134  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/276,531  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/079,677  
FILING DATE: March 27, 1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

NAME:	Lynn E. Murry, Ph.D.
REGISTRATION NUMBER:	42,918
REFERENCE/DOCKET NUMBER:	PA-0008 US
TELECOMMUNICATION INFORMATION:	
TELEPHONE:	(650) 855-0555
TELEFAX:	(650) 845-4166
INFORMATION FOR SEQ. ID NO:	47:
SEQUENCE CHARACTERISTICS:	
LENGTH:	2077 base pairs
TYPE:	nucleic acid
STRANDEDNESS:	single
TOPOLOGY:	linear
IMMEDIATE SOURCE:	
LIBRARY:	TESTTUT02
CLONE:	1271435
US-09-276-531-47	
Alignment Scores:	
Pred. No.:	5,35e-75
Score:	862.00
Percent Similarity:	87.86%
Best Local Similarity:	83.98%
Query Match:	10,78%
DB:	
US-09-720-086-6 (1-859) x US-09-276-531-47 (1-2077)	
QY	654 GlycerProCyAsnAspLeuSerAsnValAsnProAlaArgValGlyLeuTyrGluGly 673
DB	3 GGCGCGNANTGCAACGATCTTNA--AATGTGAATCA--GCCAGGAAGGCGTATGAGGAC 60
QY	674 ThrGlyArgLeuPhePheGluPheTyrHisLeuAsnTyrThrArgProGlyGly 693
DB	61 AGG---CGGCTCTTCTTGCAATTTTC-CACCTGCTGAATTACTCAGCCCAAGAGG 116
QY	694 AspAsnArgProPhePheTyrPheGluAsnValAlaMetLeuValAsnAspLeu 713
DB	117 GATGACCGGCGTCTTCTTGATGATGTTGAGAAATGTTGAGNCTCGTCCGATTCGCGCA 176
QY	714 LysAspLeuSerArgPheLeuValCysAsnProValMetLeuAspAlaLeuValSer 733
DB	177 GGGGACATCTCAGGTTCTTGAGTGTATCCAGTATGATGATGATGATGATGATGATGAT 236
QY	734 AlaAlaHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMetAsnArgProValMet 753
DB	237 GCTGCTCACAAGGCGCCGATACCTTCTTGCGGCACTACCGGATGATGATGATGATGATGAT 296
QY	754 AlaSerLysAsnAspLeuGluLeuGluAspCysLeuGluPheSerArgThrAlaLys 773
DB	297 GCATCAAAAGATATTAATCTGNGCTGCGAGCTGCTTGCAATATCAATATGATAGCCAA 356
QY	774 LeuLysLysValGlnThrIleThrThrLysSerAsnSerIleArgGlnGlyLysAsnGln 793
DB	357 TTAAGAAAGTACAGACATATACCAACCAAGTCAACTCGATCAACACAGGGGAAAAACCAA 416
QY	794 LeuPheProValValMetAsnGlyLysAspAspValLeuTyrCysThrGluLeuGluArg 813
DB	417 CTTTTCCTGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 476
QY	814 IlePheGlyPheProAlaHisIleTyrThrAspValSerAsnMetGlyArgGlyAlaArgGln 833
DB	477 NTCTTGGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 536
QY	834 LysLeuLeuGlyArgSerTyrPheSerValProValIleArgHisLeuPheAlaProLeuLys 853
DB	537 AAGCTGCTGGGAAAGGCTGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 596
QY	854 AspTyrPheAlaCysGlu 859
DB	597 GACTACTTGCATGCA 614
RESULT 2	
US-08-913-832A-1	



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 24, 2002, 12:10:37 ; Search time 68.3406 Seconds  
(without alignments)  
3854.741 Million cell updates/sec

Title: US-09-720-086-6  
Perfect score: 4590  
Sequence: 1 MKGSRHNEEBGASGYEC.....WSVPVIRHLFAPLKDYFACE 859

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
MODEL=frame+ p2n.model -DEV=xjh  
-O=/cg2\_1/uspro.spool/us09720086/runat.18112002.092957.29457/app.query.fasta.1.4252  
-DB=Issued\_Patents\_NA -OPMT=fastcap -SUFFIX=tnl -MINMATCH=0.1 -LOOPECL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09720086 @CGN 1.1.108 @runat.18112002.092957.29457 -NCTP=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_NA: \*  
1: /cg2\_6/pctdata/1/ina/5A.COMB.seq: \*  
2: /cg2\_6/pctdata/1/ina/5B.COMB.seq: \*  
3: /cg2\_6/pctdata/1/ina/6A.COMB.seq: \*  
4: /cg2\_6/pctdata/1/ina/6B.COMB.seq: \*  
5: /cg2\_6/pctdata/1/ina/PCTUS.COMB.seq: \*  
6: /cg2\_6/pctdata/1/ina/Backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	862	18.8	2077	4	US-09-276-531-47 Sequence 47, Appl
2	165	3.6	6328	4	US-08-913-832A-1 Sequence 1, Appl
3	165	3.6	6328	4	US-08-913-832A-1 Sequence 1, Appl
4	156	3.4	2355	4	US-09-249-181A-1 Sequence 12, Appl
5	150	3.3	2376	2	US-08-760-745-4 Sequence 4, Appl
6	144.5	3.1	6718	2	US-08-962-284-1 Sequence 1, Appl
7	132.5	2.9	6755	3	US-08-931-999-4 Sequence 4, Appl
8	127.5	2.8	1820	1	US-08-173-508-7 Sequence 7, Appl
9	127.5	2.8	1821	3	US-08-265-310-7 Sequence 7, Appl
10	127.5	2.8	1821	3	US-08-951-742-7 Sequence 1, Appl
11	127	2.8	2623	2	US-08-973-675-1 Sequence 1, Appl
12	125.5	2.7	34063	4	US-09-453-702B-96 Sequence 96, Appl

13	125.5	2.7	441529	4	US-09-103-840A-1	Sequence 1, Appl
14	125	2.7	4657	3	US-09-254-325-1	Sequence 1, Appl
15	125	2.7	5643	3	US-09-079-415-5	Sequence 5, Appl
16	125	2.7	5643	3	US-08-750-458A-1	Sequence 1, Appl
17	124	2.7	4181	4	US-09-393-569-1	Sequence 1, Appl
18	123.5	2.7	28804	2	US-08-592-874-1	Sequence 2, Appl
19	123.5	2.7	28804	3	US-09-096-942-2	Sequence 2, Appl
20	123.5	2.7	28804	3	US-09-096-867-2	Sequence 2, Appl
21	123.5	2.7	4403765	4	US-09-103-840A-2	Sequence 2, Appl
22	123	2.7	3044	4	US-09-152-060-36	Sequence 36, Appl
23	123	2.7	36519	3	US-08-923-137-2	Sequence 2, Appl
24	122	2.7	8931	3	US-09-051-019-1	Sequence 1, Appl
25	121.5	2.6	5965	4	US-09-362-336A-1	Sequence 1, Appl
26	121.5	2.6	6028	4	US-09-362-336A-3	Sequence 3, Appl
27	121.5	2.6	20235	1	US-07-642-734C-3	Sequence 3, Appl
28	121.5	2.6	20235	3	US-08-439-009A-3	Sequence 3, Appl
29	120.5	2.6	1170	3	US-09-307-621-1	Sequence 1, Appl
30	119	2.6	11471	4	US-09-504-358-16	Sequence 16, Appl
31	119	2.6	11471	4	US-09-954-314-16	Sequence 16, Appl
32	118.5	2.6	1865	4	US-09-370-253-5	Sequence 5, Appl
33	118.5	2.6	6463	2	US-08-962-284-3	Sequence 3, Appl
34	118	2.6	1970	1	US-08-028-463-14	Sequence 14, Appl
35	118	2.6	1970	1	US-08-461-836-14	Sequence 14, Appl
36	118	2.6	43280	2	US-08-804-227C-1	Sequence 1, Appl
37	117.5	2.6	11604	4	US-09-385-028-13	Sequence 13, Appl
38	117.5	2.6	15079	4	US-09-385-028-13	Sequence 1, Appl
39	117	2.5	2905	4	US-08-942-686-1	Sequence 1, Appl
40	117	2.5	4265	4	US-09-061-709-1	Sequence 1, Appl
41	116.5	2.5	43280	2	US-08-804-227C-1	Sequence 1, Appl
42	116	2.5	49377	1	US-08-764-233A-1	Sequence 1, Appl
43	115.5	2.5	1800	1	US-08-139-937-11	Sequence 11, Appl
44	115.5	2.5	1800	5	PCT-US93-11310-11	Sequence 11, Appl
45	115	2.5	50937	4	US-09-428-517-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1  
US-09-276-531-47  
; Sequence 47, Application US/09276531  
; Patent No. 6183968  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Lal, Preeti  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Yue, Henry  
; APPLICANT: Reddy, Roopa  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Baughn, Mariah R.  
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
; TITLE OF INVENTION: RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION  
; NUMBER OF SEQUENCES: 134  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/276,531  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/079,677  
; FILING DATE: March 27, 1998  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:



```
/ NAME: Lynn E. Murry, Ph.D.
/ REGISTRATION NUMBER: 42,918
/ REFERENCE/DOCKET NUMBER: PA-0008 US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (650) 855-0555
/ TELEFAX: (650) 845-4166
/ INFORMATION FOR SEQ ID NO: 47:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2077 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ LIBRARY: TESTTUT02
/ CLONE: 1271435
/ US-09-276-531-47
```

```
Alignment Scores:
Pred. No.: 5,35e-75 Length: 2077
Score: 862.00 Matches: 173
Percent Similarity: 87.86% Conservative: 8
Best Local Similarity: 83.98% Mismatches: 24
Query Match: 18.78% Indels: 4
DB: 4 Gaps: 1
```

US-09-720-086-6 (1-859) x US-09-276-531-47 (1-2077)

```
QY 654 GlySerProCysAsnAspLeuSerAsnValAsnProAlaArgLysGlyLeuTyrGluGly 673
DB 3 GCGCGCNANTGCAACGATCTNNA-AATGTGAATCA-GCCAGGAAAGGCTGTATAGGGAC 60
QY 674 ThrGlyArgLeuPheGluPheTyrHisLeuLeuAsnTyrThrArgProLysGluGly 693
DB 61 AGG--CGGCTCTTCGAATTTTC-CACCTGCTGAATTAACACACCCCAAGAGGGT 116
QY 694 AspAsnArgProPheMetPheGluAsnValValAlaMetLysValAsnAspLys 713
DB 117 GATGACCGGNCGTTCTTCGGAGTTTGAGAAATGTTGAGTCTGCTGCCGATTCGGGAGA 176
QY 714 LysAspIleSerArgPheLeuAlaCysAsnProValMetIleAspAlaIleLysValSer 733
DB 177 GGGGACATCTCAGCGTCTCTGGAGTGTAAATCCAGTGATGATGATGATGATGATGAT 236
QY 734 AlaAlaHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMetAsnArgProValMet 753
DB 237 GCTGCTCACAGGCGCGATCTCTGGGCAACCTACCCGGGATGAACAGGCCCGTGATA 296
QY 754 AlaSerLysAsnAspLysLeuGluLeuGlnAspCysLeuGluPheSerArgThrAlaLys 773
DB 297 GCATCAAGAATGATAAATCGNGCTGCAGGACTGCTTGGAAATACAAATAGGATAGCCAAG 356
QY 774 LeuLysLysValGlnThrIleThrLysSerAsnSerIleArgGlnGlyLysAsnGln 793
DB 357 TTAAGAAAGTACACACAATAACCAAGTCGAACCTCGATCAAAACAGGGGMAAAACCA 416
QY 794 LeuPheProValValMetAsnGlyLysAspAspValLeuTyrCysThrGluLeuGluArg 813
DB 417 CTTTTCCTCTGTTGTCATGAATGGCAAGAGATGTTTNGTGGTGCACCTGAGCTCGAAGG 476
QY 814 IlePheGlyPheProAlaHisTyrThrAspValSerAsnMetGlyArgGlyAlaArgGln 833
DB 477 NTCITTTGGCTTCTCTGCACTACACAGACGTGCTCAACATGGCCGCTGGTGGCCGCA 536
QY 834 LysLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAlaProLeuLys 853
DB 537 AAGTGTCTGGAGAGTCTCGAGCGTGCCTGTGTCATCCGACACCTCTTCGCCCCCTCT 596
QY 854 AspTyrPheAlaCysGlu 859
DB 597 GACTACTTTGCATGTGAA 614
```

RESULT 2

US-08-913-832A-1

```
/ Sequence 1, Application US/08913832A
/ Patent No. 6329517
/ GENERAL INFORMATION:
/ APPLICANT: Seelig, Hans Peter
/ APPLICANT: Renz, Manfred
/ TITLE OF INVENTION: DERMATOMYCOSITIS-SPECIFIC AUTO-ANTIGEN
/ FILE REFERENCE: 8484-0030-999
/ CURRENT APPLICATION NUMBER: US/08/913, 832A
/ CURRENT FILING DATE: 1998-01-12
/ PRIOR APPLICATION NUMBER: PCT/DE96/00444
/ PRIOR FILING DATE: 1996-03-08
/ NUMBER OF SEQ ID NOS: 2
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 1
/ LENGTH: 6328
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)...(5736)
/ US-08-913-832A-1
```

```
Alignment Scores:
Pred. No.: 2,19e-05 Length: 6328
Score: 165.00 Matches: 122
Percent Similarity: 33.90% Conservative: 79
Best Local Similarity: 20.57% Mismatches: 191
Query Match: 3.59% Indels: 202
DB: 4 Gaps: 28
```

US-09-720-086-6 (1-859) x US-08-913-832A-1 (1-6328)

```
QY 9 AsnGluGluGluGlyAlaSerGlyTyrGluGluCysIleIleValAsnGlyAsnPheSer 28
DB 43 ACTGAGGAGGAGATATGGATGCA-----CTTTGAACAACAGCTGCCCC 87
QY 29 AspGlnSerSerAspThrLysAspAlaProSerProValLeuGluAlaIleCysThr 48
DB 88 CCACCCCAACCCCAAAATGAAGAGGCCACAGAGAGGATTTGTGAGAA-----ACA 138
QY 49 GluProValCysThrProGluThrArgGlyArgArgSer-----61
DB 139 CAG-----ACTCCAAAGCTCAAGAGAAGAAAGAGCTTAAGAAACCTCGGACCCCT 189
QY 62 -----SerSerArgLeuSerLysArgGluValSerSerLeuLeuAsnTyrThrGln 78
DB 190 AAAATCCCTAAGAGCAAGCGCCAAAGAGGCGGTATGCTTTATGCGGCGAGCTGGGG 249
QY 79 AspMetThrGlyAspGly-----AspArgAspAspGluVal-----Asp 91
DB 250 GACAGCTCTGGGAGGAGGCCAGAGTTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGG 309
QY 92 AspGlyAsnGlySerAspIleLeuMetProLysLeuThrArgGluThrLysAspThrAr 111
DB 310 GACAGTGAGGCGACGACTATCTCTGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAG 369
QY 111 gThrArgSerGluSerProAlaValArgThrArgHisSerAsnGlyThrSerSerLeuGl 131
DB 370 A-----AA 372
QY 131 uArgGlnArgAlaSerProArgIleThrArgGlyArgGlnGlyArgHisHieValGlnGl 151
DB 373 GAGAAGAAGAGCAAAAT-----388
QY 151 uTyrProValGluPheProAlaThrArgSerArgArgArgAlaSerSerSerAlaSe 171
DB 389 -----CCAGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 415
QY 171 rThrProTyrSerSerProAlaSerValAspPheMetGluGluValThrProLysSerVa 191
DB 416 -----ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 450
QY 191 lSerThrProSerValAspLeuSerGlnAspGlyAspGlnGluGlyMetAspThrThrGl 211
```

Db 451 -----TTGCTCAGCTCCCTGGAAAGACTGGGCATGGAAGACATTGAC-----CA 494  
Qy 211 nvalaspalagiusserargspglaspserthrjutyrcinaspplysgluphegl 231  
Db 495 CGTCTTCAGAGGAGATTATCGAACCTTCACCACTAC----- 534  
Qy 231 ylleglasphevaltrpgllysllelysglypsererptrprcoalaetvalva 251  
Db 535 -----AAGCCTTCAGCCAGTTTGTAGACCCCTCAT 566  
Qy 251 lsertrpys-----alatrserlysaraglinalametrpoglymetartr 267  
Db 567 TGCTGCCAAATCCAGAGATTGCTGTCTCCAAGATGATGATGTTGGGTCCAAATG 626  
Qy 267 pvalglntprpheglaspgllysphe-----sergluileseralas 282  
Db 27 GCGGAGTTCAGTACCAATACCCCTTCAAGGACGTTGGGCATGCTGCAGCTGC 686  
Qy 282 plysleuvalalaleu glyleupheserghisphesleualathrphesnlysl 302  
Db 687 GGCAGCAGCAGCGGTAGCTGTGGTGAG-----AGCATGGGACAGCCACTGAGT 737  
Qy 302 uvalserlyrarlyalalawetlyrhlethrleuglulysalargvalarglaly 322  
Db 738 TGCAACCAACCTCCCTCGTGGAGGTGCTATCCGCAAGGCCAAGACCAAGAGGCGCA 797  
Qy 322 sthrpheserSerProgllygluserleugluaspglneulysPrometleuglutr 342  
Db 798 A-----GTCCTCAATGCTCGAGGAAGCCAGGAGCCCTCGTGTACTGA 845  
Qy 342 palahisclglyphelysProthrlylleuglyleulysProasnllyslglnpr 362  
Db 846 TGCCAAG-----AAGCT-----AAACCAAGAAATAGCTCC 878  
Qy 362 ovalvalasnlyl-----SerlysalarghserarSpserarshanle 377  
Db 879 CCTGAAATCAAGCTGGAGGTTTGTTCCAAGCGTAAGAGATCTCTCAGTAGCAT-- 936  
Qy 377 ugluproargargglusnlysserargargthrthrasnaspseralalase 397  
Db 937 -----GATGACTTAGATGTGAACTGACTTCGATGAGCCAGT----- 975  
Qy 397 rgluserProProlylsargleulysThrhanSerTrglylylvsasparglygl 417  
Db 76 -----ATCAATAGCTATCTGTTTCATGATGTTCCAC 1007  
Qy 417 uaspglugluserarggluargmeralasergluvalthrhanasnlyslglnleugl 437  
Db 1008 CAGCCGTAGTAGCGCAGCGCAAGAACTCCGAACCACTAAAGAAAGAAAGAGCGA 1067  
Qy 437 uaspargcyluserSerCysglylvslyshanProvalSerPhehisProleuPheglu 457  
Db 1068 GGAGGAGGTGACTGCTGTG----- 1086  
Qy 457 yglyleucysglnserCysargaspargPheleugluLeuPheTyrmetyraspgluas 477  
Db 1087 -----GA 1088  
Qy 477 pgllytr-----glnserTyrcystrhrvalCysCysgluglylYarglule 493  
Db 1089 TGGTTATGAGACAGACACCACTATATGCGAGGTGTGCGCAAGGCGGTGATGAT 1148  
Qy 493 uleucysserasenthserCysargCyspheCysvalglucysleugluvalleuva 513  
Db 1149 CCTGTGTGAT-----ACCTGTCCCTGCTTACCAATGCTGCTGCTGAT----- 1194  
Qy 513 lgllyalaglythrleuaspalalyseuglnlyl--ProtpserCysTyrmetyr 532  
Db 1195 -----CCCACATGAGAGAGCTCCGAGGCGCAAGTGGAGCTGCCACACTG 1241  
Qy 532 sleuProglinargCyshtisgllyvalleuarargarglylsaprtprasnmetargleugl 552  
Db 532 ----- 552

Db 1242 CGAAGAGAA-----GGCATCCAGTGGAGGCTAAAGAGACAAATTCGAGGGTGA 1292  
Qy 552 naapPhephe---ThrThraspProapleuglu 563  
Db 1293 GGAGATCCTCGAAGAGGTTGGGGAGAGACTCGAAGAG 1329

RESULT 3  
US-09-249-181A-1  
; Sequence 1, Application us/09249181A  
; Patent No. 6440679  
; GENERAL INFORMATION:  
; APPLICANT: Seelig, Hans Peter  
; APPLICANT: Renz, Manfred  
; TITLE OF INVENTION: DERMATOMYOSITIS-SPECIFIC AUTO-ANTIGEN  
; FILE REFERENCE: 8484-0059-999  
; CURRENT APPLICATION NUMBER: US/09/249,181A  
; PRIOR FILING DATE: 1998-02-12  
; PRIOR APPLICATION NUMBER: US 08/913,832  
; PRIOR FILING DATE: 1998-01-12  
; PRIOR APPLICATION NUMBER: PCT/DE96/00444  
; PRIOR FILING DATE: 1996-03-08  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 6328  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)...(5736)  
US-09-249-181A-1

Alignment Scores:  
Pred. No.: 2,19e-05 Length: 6328  
Score: 165.00 Matches: 122  
Percent Similarity: 33.908 Conservative: 79  
Best Local Similarity: 20.574 Mismatches: 191  
Query Match: 3,594 Indels: 202  
Gaps: 28

US-09-720-086-6 (1-859) x US-09-249-181A-1 (1-6328)

Qy 9 Asnlguglucgllyalaserglytrgluclycyllelevalasnlylvanpneser 28  
Db 43 AGTGAAGAGAGATATGATGCA-----CTTTGAACAACAGCCCTGCC 87  
Qy 29 AspqlnserSeraspthrlyaspalaproserProProvalleuglualeCysThr 48  
Db 88 CCAACCCACCCGAAATGAAGAGAGCCAGAAAGAGATTTGTCAAA-----ACA 138  
Qy 49 GluprovalCysThrProglutThrArglylYargSer----- 61  
Db 139 GAG-----ACTCCAAAGCTCAAGAAAGAAAGAAAGCCTTAAGAAACCTGGACCT 189  
Qy 62 -----SerSerargleuserlysargluvalserSerleuLeuasnlyrThrln 78  
Db 190 AAAATCCCTTAAGAGCAAGCCCAAAAAAGAGCGTATGCTTATGCGGAGCTGGGG 249  
Qy 79 Aspmethrlyaspglly-----Aspargaspaspgluval-----Asp 91  
Db 250 GACAGCTCTGGAGGAGGCGCAGAGATTGTGTGAGAGAGAGAGAGAGGTGCTTGCCTCA 309  
Qy 92 AspqllyasnlySerasp-IlleuMetProlylsleuThrArglylThrlyAspThrAr 111  
Db 310 GACAGTGAAGGAGCAGCATATCTCTGCGCAAGAGAGAGAGAGAGAGAGAGAGAG 369  
Qy 111 gThrArgsergluserProalavalargThrArghtisSerAsnlylThrSerSerleugl 131  
Db 370 A-----AA 372  
Qy 131 uarglunarglaserProarglileThrArglylYargGlnlylYarghtisIleValGlnl 151  
Db 373 GAGAAAGAGAGCAAAAT----- 388

QY 493 uLeuCySerAenThrSerCysArgCysPheCysValGluCysLeuGluValLeuVa 513  
 Db 1149 CTTGTGTGAT-----ACCTGTCCCCGTGCTTACCACATGCTGCTCTGGAT----- 1194  
 QY 513 lGlyAlaGlyThrAlaGluAspAlaLysLeuGlnGlu---ProTrpSerCysAtyMetCy 532  
 Db 1195 -----CCCGAGATCGAGAGAGCTCCCGAGGCGCAAGTGGAGCTGCCACACTG 1241  
 QY 532 sLeuProGlnArgCysHisGlyValLeuArgArgArgLysAspTrpAlenMetArgLeuGl 552  
 Db 1242 CGAGAAGGAA-----GGCATCCAGTGGGAAGCTTAAGAGAGACCAATTCGGAGGGTGA 1292  
 QY 552 nAspPhePhe---ThrThrAspProAspLeuGluGlu 563  
 Db 1293 GGAGATCTCGAAGAGGTTGGGGGAGACCTCGAAGAG 1329  
 RESULT 4  
 US-08-913-159-12  
 ; Sequence 12, Application US/08913159  
 ; Patent No. 6300109  
 ; GENERAL INFORMATION:  
 ; APPLICANT:  
 ; TITLE OF INVENTION: Plasmid-derived type II  
 ; TITLE OF INVENTION: restriction-modification systems from *Lactococcus lactis*  
 ; NUMBER OF SEQUENCES: 14  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/913,159  
 ; FILING DATE:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: DK 0179/95  
 ; FILING DATE: 17-FEB-1995  
 ; INFORMATION FOR SEQ ID NO: 12:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2355 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; HYPOTHETICAL: NO  
 ; ANTI-SENSE: NO  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: *Lactococcus lactis* subsp. *cremoris*  
 ; STRAIN: W39  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 744..1283  
 ; IDENTIFICATION METHOD: experimental  
 ; OTHER INFORMATION: /codon\_start= 744  
 ; OTHER INFORMATION: /product= "LlaDII restriction endonuclease"  
 ; OTHER INFORMATION: /evidence= EXPERIMENTAL  
 ; OTHER INFORMATION: /gene= "ORF"  
 ; OTHER INFORMATION: /number= 1  
 ; OTHER INFORMATION: /standard\_name= "Gene coding for R.LlaDII"  
 ; OTHER INFORMATION: /label= r-lladII  
 ; OTHER INFORMATION: /note= "The first ten amino acids in this sequence may be  
 ; OTHER INFORMATION: doubtful. However, from base 773 this reading frame gives  
 ; OTHER INFORMATION: homology with the Bsp61 endonuclease"  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1392..2342  
 ; IDENTIFICATION METHOD: experimental  
 ; OTHER INFORMATION: /codon\_start= 1392  
 ; OTHER INFORMATION: /product= "LlaDII methylase"  
 ; OTHER INFORMATION: /evidence= EXPERIMENTAL  
 ; OTHER INFORMATION: /gene= "ORF"  
 ; OTHER INFORMATION: /number= 2  
 ; OTHER INFORMATION: /standard\_name= "Gene coding for M.LlaDII"  
 ; OTHER INFORMATION: /label= m-lladII

OTHER INFORMATION: /note= "The sequence shows 60 % identity and 76 % similarity  
; OTHER INFORMATION: with the Bsp61 methylase."  
US-08-913-159-12

## Alignment Scores:

Pred. No.:	3,18e-05	Length:	2355
Score:	156.00	Matches:	78
Percent Similarity:	35.48%	Conservative:	43
Best Local Similarity:	22.87%	Mismatches:	124
Query Match:	3.40%	Indels:	96
DB:	4	Gaps:	16

US-09-720-086-6 (1-859) x US-08-913-159-12 (1-2355)

```
OY 581 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeu 600
DB 1395 TTGAATAATGCTTCTTTTTCGCCGAGTTGGCGAATTGATTGTTTGAATAATGCA 1454
OY 601 GlyIleLysValGluLysTyrIleAlaSerGluValCysAlaGluSerIleAlaValGly 620
DB 1455 GGTTCACAAACA-----ATATATCTAATGAATTT-----GATATATATGCTGCTGAT 1502
OY 621 ThrValLysHisGluGlyGlnIleLysTyrValAsnAspValArgLysIleThrLysLys 640
DB 1503 ACTTTTGAATGAACCTTGAAGTAAAGTA-----GACCGACGATATATAATGATGTA 1556
OY 641 AsnIleGluIleTyrGlyProPheAspLeuValIleGlyLysSerProCysAsnAspLeu 660
DB 1557 CAAAGCTGATGAATATACAGATTTGATATATGATGATGATGATGATGATGATGATGAT 1616
OY 661 SerAsnValAsnProAlaArgLysGlyLeu-----TyrGluGlyThrGlyArgLeuPhe 678
DB 1617 TCT--ATTGCTGGTATTCGTCACAGCCTTACAGATGAACAAGTCGAGTAACTTTT 1673
OY 679 PheGluPheTyrHisLeuLeuAsnTyrThrArgProLysGluGlyAspAsnArgProPhe 698
DB 1674 TTTGAACCTTGTCTGATTTTGAACCAAAAAAACCTCGTGTGCA----- 1718
OY 699 PheTyrPhePheGluAsnValValAlaMetLysValAsnAspLysAspIleSerArg 718
DB 1719 -----TCTTTGAAAAATGTTAAAAATCTGTTTTCACGATAGCGGAAACACATTTGA 1772
OY 719 PheLeuAlaCysAsnProValMetIleAspAlaIleLysValSerAlaIleArgAla 738
DB 1773 GTTATTGTG-----TCTGAGTTAGAAAGACTA 1799
OY 739 ArgTyr-----PheTyrGlyAsnLeuProGlyMet 748
DB 1800 GGGTACAAAGTATCTTTTCAAGGTGTTAAATGCTTGAATATGAAATATATACCT---CAA 1856
OY 749 AsnArgPro-----ValMetLysSerLysAsnAsp----- 758
DB 1857 AATAGAGAAAGTATCTATATATTGCTTCAAAAAATAAAAAGATTATGCAAAATTTGAA 1916
OY 759 -----LysLeuGluLeuGlnAspCysLeuGluPheSerArg--- 770
DB 1917 CTACCAAAATCTATACCTTTAAAAACACAGATTGATGATGATGATGATGATGATGAT 1976
OY 770 ----- 770
DB 1977 CAAAGCAGATAAGTTCTACTATACCTCTGAAAAGATAAATTTTGTATGAGTTAAAAAGAA 2036
OY 771 -----ThrAlaLysLeuLysLysValGlnThrIleThrThrLys 783
DB 2037 AATATGACTAATACAGACATACATATACAGTGGCTGAGATTATATGTAAGAAACAAACAA 2096
OY 784 SerAsnSerIle-----ArgGlnGlyLysAsnGlnLeuPhe 795
DB 2097 AGTAAATTTAGTACCAACACTAACGCGCTAATATGGAACAGGTGGGCAATAAT-----GTG 2150
OY 796 ProValValMetAsnGlyLysAspAspValLeuTyrCysThrGluLeuGlu-----Arg 813
DB 2151 CCTATATCTTACATATATAGCGGAGATATTCGTAATTAACACCAAGCAATATGCTTTAAC 2210
```

```
OY 814 IlePheGlyPheProAlaHisTyrThrAspValSerAsnMetGlyArgGlyAlaArgGln 833
DB 2211 GTTCAAGGTTTCCCAAAAGATAT---AAACTTCAAAACCAAGTAAATGGAGATTATAT 2267
OY 834 LysLeuLeuGlyArgSerTyrPheValProValIleArgHisLeuPheAlaProLeuLys 853
DB 2268 AAACAGACAGAAACAGTGTGTAGTACCACTATATGAAAGAAATGCAAAAAATCTTGCA 2327
OY 854 Asp 854
DB 2328 GAT 2330
```

## RESULT 5

US-08-760-745-4

; Sequence 4, Application US/08760745

; Patent No. 5972658

; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Goli, Surya K.

; APPLICANT: Murry, Lynn E.

; TITLE OF INVENTION: LUNG GROWTH FACTOR VARIANT

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: US

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/760,745

; FILING DATE: Herewith

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0169 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; TELEX:

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2376 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: GenBank

; CLONE: 598956

; US-08-760-745-4

## Alignment Scores:

Pred. No.:	0.000126	Length:	2376
Score:	150.00	Matches:	83
Percent Similarity:	36.62%	Conservative:	58
Best Local Similarity:	21.56%	Mismatches:	163
Query Match:	3.27%	Indels:	81
DB:	2	Gaps:	15

US-09-720-086-6 (1-859) x US-08-760-745-4 (1-2376)

```
OY 142 GlyArgGlnGlyArgHisValGlnGlyTyrProValGluPheProAlaThrArgSer 161
DB 142 ----- 161
```

Db 25 GGTGGAGGAAGAGCGCTCCGCGAGAGGAGGAGCAATTAATTTCAAAACACAACT 84  
Qy 162 ArgArgArgAlaSerSerAlaSerThrProTrpSerSerProAlaSerValAsp 181  
Db 85 CGACGAGCGCGACCCACCGCGCGGAGCTTCCCGATCCGCGCGCGCC-----CCG 138  
Qy 182 PheMetGluGluValThrProLysSerValSerThrProSerValAspLeuSerGlnAsp 201  
Db 139 TCCGTGCGCGCGCGCGGAGACCGCGCGCGCGCGCGCGCGCGCGCTC----- 183  
Qy 202 GlyAspGlnGluGlyMetAspThrThrGlnValAspAlaGluSerArgAspGlyAspSer 221  
Db 184 GGGCC 243  
Qy 222 ThrGluThrGlnAspAsp----- 227  
Db 244 ACCGCC 303  
Qy 228 -----LysGluPheGlyIleGlyAspLeuVal 236  
Db 304 CGGAGCG 363  
Qy 237 TrpGlyLysIleLysGlyPheSerTrpTrpProAlaMetVal-----Val 251  
Db 364 TTCGCAAGATGAAGGCTACCCACACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 423  
Qy 252 SerTrpLysAlaThrSerLysArgGlnAlaMetProGlyMetArgTrpValGlnTrpPhe 271  
Db 424 GCCGTGAATCAACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 465  
Qy 272 GlyAspGlyLysPheSerGluIleSerAlaAspLysLeuValAlaLeuGlyLeuPheSer 291  
Db 466 GGGACCCAGAGCGCGCATCTCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 525  
Qy 292 GlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTrpArgLysAlaMetTyHis 311  
Db 526 GAGAAGTTTGGCAAGCC-----AACAAAGGAGAAAGGTTACGCGGCGCGCGCGG 579  
Qy 312 ThrLeuGluLysAlaArgValArgAlaGlyLysThrPheSerSerProGlyGluSer 331  
Db 580 ATCAGAGAACACCTTACTGTCAAGGCT--TCCGGCTATCAGTCTCTCCAGAAAAGAGC 636  
Qy 332 LeuGluAspGlnLeuLysProMetLeuGluTrpAlaHisGly---GlyPheLysProThr 350  
Db 637 TGTGTGAAGAGCCTGAACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 696  
Qy 351 GlyIleGluGlyLeuLysProAsnLysLysGlnProValValAsn----- 365  
Db 697 AATGACAGAGCG 756  
Qy 366 -----LysSerLysValArgArgSerAspSerArgAsnLeuGlu----- 378  
Db 757 AAGAACGAGAGGAGCGGTTGAAGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 816  
Qy 379 ---ProArgArgGluAsnLysSerArgArgThrThrAsnAspSerAlaAlaSer 397  
Db 817 CGTCCCAAGAGCGGT 876  
Qy 398 GluSerProProLysArgLeuLysThrAsn-----SerTyArgLysLysAsp 414  
Db 877 GAGAGGCGCGCTTCTATGAGGTGGAAAGAAATAGCACCCCTCTGAGCGCGCGCGCT 936  
Qy 415 ArgGly-----GluAspGluGluSerArgArgMetAlaSerGlu 428  
Db 937 CGGGGCGCTCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 996  
Qy 429 ValThrAsnAsnLysGlyAsnLeuGluAsp-ArgCysLeuSer-----CysGlyLysLy 446  
Db 997 GATCTGAGCG 1056  
Qy 446 sAsnProValSerPheHisPro-----LeuPheGluGlyLysLeuCy 460  
Db 1057 GAGCG 1107

Qy 460 sGlnSerCysArg 464  
Db 1108 CATGGCGCTCAAA 1120

RESULT 6  
US-08-962-284-1  
; Sequence 1, Application US/08962284  
; Patent No. 5985608  
; GENERAL INFORMATION:  
; APPLICANT: Luna, Elizabeth J.  
; APPLICANT: Pestonjamas, Kersi N.  
; APPLICANT: Pope, Robert K.  
; APPLICANT: Wulfschuh, Julia D.  
; TITLE OF INVENTION: ACTIN-BINDING POLYPEPTIDES  
; TITLE OF INVENTION: AND NUCLEIC ACIDS ENCODING THE SAME  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/962,284  
; FILING DATE: 31-OCT-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fasse, Peter J.  
; REGISTRATION NUMBER: 32,983  
; REFERENCE/DOCKET NUMBER: 07917/058001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 6718 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: Coding Sequence  
; LOCATION: 451...5814  
US-08-962-284-1

Alignment Scores:  
Pred. No.: 0.00256 Length: 6718  
Score: 144.50 Matches: 217  
Percent Similarity: 32.24% Conservative: 126  
Best Local Similarity: 20.39% Mismatches: 370  
Query Match: 3.15% Indels: 352  
DB: 2 Gaps: 55

US-09-720-086-6 (1-859) x US-08-962-284-1 (1-6718)

Qy 6 ArgHisLeuAsnGluGluGluGlyAlaSerGlyTyArgGluCys----- 20  
Db 1322 AGGTGGATGAACGAGCGCAAGCTGAGCGTCCCGCCCAAGAGGTTGCTTTTCAGGGAGATGG 1381  
Qy 21 -----IleIleValAsnGlyAsnPheSerAspGlnSerSerAspThrLys----- 35  
Db 1382 AAAAATCTTTTGATGAACAAATGTTCCAAAGCGACCTCAAGAAACACAGCTGTGGAGC 1441  
Qy 36 -----AspAlaProSerProProValLeuGluAlaIle 46  
Db 1442 AGAGGCTACGCCGTCTGCGAGGACAGGTCCTCACCACCGCCCATCACCACCTGAAGAGGTGG 1501

47 CysThrGluProValCysThr-----ProGluThrArgGly 58  
Db 1502 TCATCGACGCCACATTGCGAGCCCTGCTGTCACCAAAAGGCTTAGAGCAAGACACAA 1561  
Gly ArgArgSerSerSerArgLeuSerLysArgGluVal-SerSerLeuLeuAntyThrG 78  
Db 1562 ATGGGGGAAAGAGCTTGCTGAGCAAGAGAACTGATTCCTCACTCACTGAGCTTGCGCC 1621  
Gly 78 InAspMetThrGlyAspGlyAspArgAspGluValAspAspGlyLysnGlySerAspI 98  
Db 1622 AAAAGTTGGCC----- 1632  
Gly 98 LeuMetProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSerPro 118  
Db 1633 --TTGTTTAAACAATTTGTCAGCAGTC-----TCAAAAG 1666  
Gly 18 LalaArgThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnArgAlaSerPro 138  
Db 1667 CGATTTCTACCGGAAACAGAAATAGACAG-----AGACAGAGAGA----- 1707  
Gly 138 rglLeThrArgGlyArgGlnGlyArgHisValGlnGluThrProValGluPhePro 158  
Db 1708 -----ATGACGCTGCTATCAAACTGAG-----CCAG 1735  
Gly 158 LathArgSerArgArgArgArgAlaSerSerSerAlaSerThrProTrpSerSerPro 178  
Db 1736 TCACACTGGAGAGCTGAGACAGGTGACAGATGGAAAGCTCATTCCTTCTCACCTGCCG 1795  
Gly 178 LaseValAspPheMetGluGluValThrProLysSerValSerThr-----ProSerV 196  
Db 1796 TGAACACAC-----TCAGTGTCTACCGTAGCATCCACGG 1828  
Gly 196 alAspLeuSerGlnAspGlyAspGlnGluGlyMetAspThrThrGlnValAspAlaGluS 216  
Db 1829 TTGCTCCCATGATGATCGCGAGATCTTCGCAACAAAGCCACTTGACCAACATGCAAGTG 1888  
Gly 216 erArgAsp-----GlyAspSerThrGluThrGlnAspAspLysGluPheGlyIleG 233  
Db 1889 CCACGTACTATAGTTTCTTCTTCAATAGAAATTCGACTCTCCAGTTAGAACATT- 1947  
Gly 233 LysAspLeuValTrpGlyLysIleLysGlyPheSerTrpTrpPro----- 247  
Db 1948 -----CTGAATGCGCAAGCTTGCGAGCCTTTGGTAGAGGTAAGC 1987  
Gly 48 -----AlaMetValIserTrpLysAlaThrSerLysArgGlnAlaMetProGlyM 265  
Db 1988 AGAACAAGGAAATGTTGAGAGATATGAGACAGACAGAAAGCAAGAGCTTGACAGGTC 2047  
Gly 265 erArgTrpValGlnTrpPheGlyAsp-----GlyLysPheSerGluI 279  
Db 2048 GA-----GACAGTGGGATGAGAAAGTATGGTCTTTGAGGAG 2086  
Gly 279 LeSerAlaAsp----- 282  
Db 2087 CAGAAAGCATCTCAACCCCATCTGAAACAGAGCCAGGAAAGACAGCCATTAGCAATCTA 2146  
Gly 283 -----LysLeuValAlaLeuG 288  
Db 2147 AATATGCTGTTCCCAAGAGAGAAAGCTGGAACGGGGAACCTCCATCACCACCTCG 2206  
Gly 288 LysLeuPheSerGlnHisPheAsnLeuAlaThrPheAsn----- 300  
Db 2207 GGGATGAAACCGAAGAAATTTTCCATGGCTTAAATGATGACACAGAAACTTGACATTGA 2266  
Gly 301 -----LysLeuValS 304  
Db 2267 GGGACAGGCTGCCCTTTGAAGAGAGAGTGGAGTGAAGATGTTATGAAGAAAGATTTT 2326  
Gly 304 erTyArgLysAlaMetTy-----HisThrLeuGluLysAlaArgValArgAlaGlyL 322  
Db 2327 CACTAAGAGCGGACAGATTCGGGGAGCCCACTTCGAGGACAGACGGGACAGCTGCTGGGA 2386

Gly 322 YThrPheSerSer-----ProGlyGluSerLeuGluAspG 335  
Db 2387 AAACCTATTGCTCAAAACACAGAGCCCGTGTCTGGAAGCCCGCAGGATTTCTCGGAACAGC 2446  
Gly 335 InLeuLys-----PrometLeuGluTrpAlaHisGlyLysPheLys---P 349  
Db 2447 CACAGAGAAAGCTTCGAAAGAAATCCATGTCGATGTTCTGCTGGAGAGATCAAAAGCC 2506  
Gly 349 roThrGlyIleGluGlyLeu-----LysProAsnLys-----LysGlnProV 363  
Db 2507 CGACAGGG--GAGGGCTTTCTTGACTCACCAGCAAAACATGCTTATTAAGAAAGAT 2563  
Gly 363 alValAsnLysSerLysValArgArgSerAspSerArgAsnLeuGluProArgArgArg 383  
Db 2564 TGGCAGCTGTTAAGAAAGAGAGAGAGAGATGGAGAAACAGACTACAGCAGAGAGCAGC 2623  
Gly 383 Lu--AsnLysSerArgArgThrThrAsnAspSerAlaAlaSerGluSerProProp 402  
Db 2624 AGGGCGGCAAGCGCGCGCAGAGCTGACACCCAGAGAGAGGGCGGCTTCATCA 2683  
Gly 402 roLysArgLeuLysThrAsnSerTyrgLysLysAspArgGlyGluAspGluGluSerA 422  
Db 2684 AGAAGCGGTC-----ACAGAAAGTCGAGAGAGCCAAATGACGA 2722  
Gly 422 rglLysArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAspArgCysLeuS 442  
Db 2723 TTGAGAGAGAGAGAGAGCTCATCACTGT-GAGAGAGAGAGCTTGAAAGACGAGAGCACA 2781  
Gly 442 erCysGlyLysLys--AsnProValSerPheHisProLeuPheGluGlyGlyLeuCysG 461  
Db 2782 GGAAGCGGCAAGACTCGACCCGATT-----CACTGTGCTGCGAGATGGT----- 2828  
Gly 461 InSerCysArgAspArgPheLeuGluLeuPheTyrmetyrAspGluAspGly----- 478  
Db 2829 -----GAGAAAGGTTT--GGCGTCACTACTGCCATTAACCCCACTAGCCTACGCC 2877  
Gly 479 -----TyrgInserT 482  
Db 2878 ATTTCGGTAAACAAAGAGCAGACACCCGTTTCCAAACCCCTGGAAGATATCGAAGCC 2937  
Gly 482 YrCysThrValCysCysGlu--GlyArgGluLeuLeuLeuCysSerAsnThrSerCysC 501  
Db 2938 AGACAGATATGACAGTTAGATGAGACTGGAAGTTGACAGAGCTGGAACCTTTCT- 2993  
Gly 501 YsArgCysPheCysValGluCysLeuGluValLeuValGlyAlaGlyThrAlaGluAspA 521  
Db 2994 -----AAGAAGCTGATTAAC 3009  
Gly 521 LalsLeuGlnGluProTrpSerCysTyrmetyrCysLeuProGlnArgCysHisGly-Val 540  
Db 3010 AAAGTTGGCGGAGTGCAGAAACGCTGCTCACTGTCACCGCAATCTGTGAAGAGGTG 3069  
Gly 541 LeuArg-----ArgArgLysAspTrpAsnMet 549  
Db 3070 ATGAAGCCAGATGATGATGAACCTTCCAAATTTTACCGCAGCGATATATATATG 3129  
Gly 550 ArgLeuGlnAspPheThrThrAspProAspLeuGluGlu--PheGluPro----- 566  
Db 3130 CCA--AGAACTCTGTGAGATGATGAGAGCTTCGATGCTATTTGCATCTTATGCA 3186  
Gly 567 ProLysLeuTyProAlaIleProAlaIleLysArg----- 578  
Db 3187 CCCAATTTGAGCTCTTCGTCGCGGCGAGACAAAGCGGCACTTAGGCCCAAGCGCGGCTT 3246  
Gly 579 -----ArgProIleArgValLeuSerLeuPheAspGlyIleAlaThrGlyTy 594  
Db 3247 CAGGCTCCAAAACCCCTGGAATGCTGGCGGCAAGAGATCTCTTCAGAGATAC 3306  
Gly 595 -----LeuValLeuLysGluLeuGlyIleLysValGluLysTy 607  
Db 3307 ACTGAGCAGAGATTAACGTTGCTTCATGAGTCAAAAGCGGATGAAGAGTGAAGAAAGATG 3366  
Gly 608 IleAlaSerGluValCysAlaGluSerIleAlaValGlyThrValLysHisGluGly--- 626





QY 158 lATHrGserGArGArGArGAlAserSerSerAlaSerThrProTrpSerSerPro 178  
 Db 4824 CGACGACCAAAAAAGAGAGAGAGACCCACGAAAGAGACACCCCCCAACCCAGAAAA 4883  
 QY 178 lAserValAspPheMetCgluIuValThrProIysSerValSerThrProSerValAspL 198  
 Db 4884 AAAAAACAAGCAAGAGAGACGAGAAAAACACAGAAAAAAGAAAGCCACGGCAAGAGAGC 4943  
 QY 198 euserGlnAspGlyAspGlnGluGlyMetAspThrThrGlnValAspAlaGluSerArgA 218  
 Db 4944 AAAAAAGAAAAAGAAAAAGAAAAAGAAAGAAAAAGAGAGA-AAAAACAACAACGG 5002  
 QY 218 spGlyAspSerThrGluTrpGlnAspAspIysGluPheGlyIleGlyAspLeuValTrpG 238  
 Db 5003 GGGGGGCAAGCAGGAGACCAAGAGCAGAGACAGAAAAA-----ACGG 5041  
 QY 238 llyleIleIysGlyPheSerTrpTrpProAlaMetValAlaSerTrpIysAlaThrSerL 258  
 Db 5042 GAAAAACAATAAAC-----AGCAAAAAAGCAAAACCA 5074  
 QY 258 lysArgGlnAlaMetProGlyMetArgTrpValGlnTrpPheGlyAspGlyIysPheSerG 278  
 Db 5075 AAGAGAACCAAAAAACAGGGCAGGCCAAAGCCGCAACACACACAAGACAAACAAGAGAGG 5134  
 QY 278 lulleSerAlaAspIysLeuValAlaIleuGlyLeuPheSerGlnHisPheAlaLeuAlaT 298  
 Db 5135 AA----- 5136  
 QY 298 hrPheAsnIysLeuValSerTrpArgGlyAlaMetTrpHisThrLeuGluIysAlaArgV 318  
 Db 5137 -----CACAAAACCGGCAAAAGGGAAACAAAAAGAGAAAGG- 5175  
 QY 318 alArgAlaGlyIysThrPheSerSerSerProGlyGluSerLeuGlnAspGlnLeuIysP 338  
 Db 5176 -----GGGGAAAAAAGAGAAAGAGGACAGGGAGGGGACAGAAAGAAAGCCGCCAA- 5226  
 QY 338 rometLeuGluTrpAlaHisGlyGlyPheIysProThrGlyIleGluGlyLeuIysProA 358  
 Db 5227 -----AAAAAGA 5233  
 QY 358 snIyelysGlnProValValaIasnIysSerIysValArgArgSerAspSerArgAsnLeuG 378  
 Db 5234 AAAAAACCGGAACCGAAAAAGCAAGAACCAAGAAAAAAGAAAGAGAGCAAAAAACGGCG 5293  
 QY 378 luProArgArgArgGluAsnIysSerArgArgArgThrThrAsnAspSerAlaAlaSerG 398  
 Db 5294 AAAAAACAGAAAAAAGAACACACAAACAAGGAGCAGAAC-----GGAC 5335  
 QY 398 luserProProIysArgIleuIysThrAsnSerTrpGlyIlyAspAspArgGlyGluA 418  
 Db 5336 AACAAAAACCGAAAAAGAGACGGAAAAACCAAAACGAACCCGAAAAAGCGAAGAGACAAG 5395  
 QY 418 spGluGluSerArgGluArgMetAlaSerGluValThrAsnAsnIys 433  
 Db 5396 CAACCAACGAAAAAGAGACAAAGCGAAAAAGAAAAAAGAAAAA 5442  
 RESULT 8  
 US-08-173-508-7  
 ; Sequence 7, Application US/08173508  
 ; Patent No. 5616485  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bartfield, Daniel  
 ; APPLICANT: Butler, Michael J.  
 ; APPLICANT: Hadary, Dany  
 ; APPLICANT: Jenish, David  
 ; APPLICANT: Krieger, Timothy  
 ; TITLE OF INVENTION: STREPTOMYCES PROTEASES AND IMPROVED  
 ; TITLE OF INVENTION: STREPTOMYCES STRAINS FOR EXPRESSION OF PEPTIDES AND  
 ; NUMBER OF SEQUENCES: 13  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Foley & Lardner  
 ; STREET: 3000 K Street, N.W.  
 ;

```

1 CITY: Washington, D.C.
2 COUNTRY: USA
3 ZIP: 20007-5109
4 COMPUTER READABLE FORM:
5 MEDIUM TYPE: Floppy disk
6 COMPUTER: IBM PC compatible
7 OPERATING SYSTEM: PC-DOS/MS-DOS
8 SOFTWARE: Patentin Release #1.0, Version #1.25
9 CURRENT APPLICATION DATA:
10 APPLICATION NUMBER: US/08/173,508
11 FILING DATE: 23-DEC-1993
12 CLASSIFICATION: 435
13 ATTORNEY/AGENT INFORMATION:
14 NAME: BENT, Stephen A.
15 REGISTRATION NUMBER: 29,768
16 REFERENCE/DOCKET NUMBER: 18740/125/CACO
17 TELECOMMUNICATION INFORMATION:
18 TELEPHONE: 202 672 5300
19 TELEFAX: 202 672 5399
20 TELEX: 904136
21 INFORMATION FOR SEQ ID NO: 7:
22 SEQUENCE CHARACTERISTICS:
23 LENGTH: 1820 base pairs
24 TYPE: nucleic acid
25 STRANDEDNESS: double
26 TOPOLOGY: linear
27 FEATURE:
28 NAME/KEY: CDS
29 LOCATION: 104..1720
30 FEATURE:
31 NAME/KEY: sig_peptide
32 LOCATION: 104..244
33 FEATURE:
34 NAME/KEY: mat_peptide
35 LOCATION: 245..1720
36 US-08-173-508-7
37
38 Alignment Scores:
39 Pred. No.: 0.0134 Length: 1820
40 Score: 127.50 Matches: 145
41 Percent Similarity: 33.95% Conservative: 75
42 Best Local Similarity: 22.38% Mismatches: 251
43 Query Match: 2.78% Indels: 182
44 DB: 1 Gaps: 30
45
46 US-09-720-0862.6 (1-859) x US-08-173-508-7 (1-1820)
47
48 QY 37 AlaProSerProProValLeuGluAlaIleCyThrGluProValCyThrPro-GluTh 56
49 |||:::|||||:::|||||:::|||||:::|||||
50 Db 135 GCACCGGGGACACCGGTTCCGGGCCACGCTGCTCACCGCCGGCTGCTGCCACCGCCT 194
51
52 QY 56 rArgGlyrArgrGserSerSerArgLeuSerLysrArgGluValSerSerLeuLeuAsnTy 76
53 |||||:::|||||:::|||||:::|||||
54 Db 195 GCTCGGCGCGG-----GGGCGCGCTGACGTCGCGCGAGCCCCCGGCCAAGCGCG 245
55
56 QY 76 rThnGlnAspMetThnGlyAspGlyAspArgAspAspGluValAspAspGlyAsnGlySe 96
57 ::::|:::|||||:::|||||:::|||||
58 Db 246 CCGGCGGACGAGGCGGCGCACCGCGACCTGACCCCCCTGCCAA-----GGCCACGC 299
59
60 QY 96 rAspIleLeuMetProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSe 116
61 ::::|:::|||||:::|||||:::|||||
62 Db 300 CCGCGGAGCTGCCCCGCTACTACGACGAGAAGCTGCGCTGCGCGCACTGCGG-----CG 353
63
64 QY 116 rProAlaValArgTrnArgHisSerAsnGlyThrSerSerLeuGluArgGlnArg----- 134
65 ::::|:::|||||:::|||||:::|||||
66 Db 354 TCCGGGGCTTCAGTGGCGGCACCATGAAGGCCCCCTCGACTACGCCAAGGCCCGCGACG 413
67
68 QY 135 -----AlaSerProArgI 139
69
70 Db 414 GCCAGCTCCGCGCTCGCGGTGGCCCGCAGAAAGCCACGAGGCGCGGCGGACGCTCGCGCT 473
71
72 QY 139 eThnArgGlyrArgGlnGlyrArgHisIshValGlnGluTyrrProValGluPheProAlaTh 159
73 :|||:::|||||

```





REFERENCE/DOCKET NUMBER: 18740/133/CACO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202 672 5300  
TELEFAX: 202 672 5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1821 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 104..1720  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 104..244  
ATTURE:  
NAME/KEY: mac\_peptide  
LOCATION: 245..1720  
US-08-265-310-7

Alignment Scores:  
Pred. No.: 0.0134 Length: 1821  
Score: 127.50 Matches: 145  
Percent Similarity: 33.95% Conservative: 75  
Best Local Similarity: 22.38% Mismatches: 251  
Query Match: 2.78% Indels: 182  
DB: Gaps: 30

US-09-720-086-6 (1-859) x US-08-265-310-7 (1-1821)

QY 37 AAlProserProProvalLeuGluAlaLeCyethrGluProvalCystrPro-GluTh 56  
Db 135 GCACCGGGGCGACCCGTTTCCGGGCGACGCTGCTCACCAGCGGCTGCTGCCACCGCT 194  
QY 56 rArgGlyArgArgSerSerArgLeuSerlyrArgGluValSerSerLeuLeuAenTy 76  
Db 195 GCTGGGCGG-----GGGCGGCTCGACGCTCCCGGATCCCGCGGCGCAAGGCGG 245  
QY 76 rThrGlnAspMetThrGlyAspGlyAspArgAspAspGluValAspAspGlyAsnGlyse 96  
Db 246 CCGGCGGCGACGAGGCGCGCCACGCGGACCTGACCCCTGCGCA-----GCCACGC 299  
QY 96 rAspIleuMetSerProlySleuThrArgGluThrLyAspThrArgThrArgSerGlyse 116  
Db 300 CCGGCGAGCTGCTCCCTACTACGACGACAGAGCTCGGCTGGCGGCACTGCGG-----CG 353  
QY 116 rProAlaValArgThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnArg----- 134  
Db 354 TCCCGGGCTTCCAGTGGCGCCACATGAGGCCCCGCTGACTACGCCAACCCCGCCACG 413  
QY 135 -----AlaserProArgI 139  
Db 414 GCAGACTCCGGCTCGGCGTGGCCCCGCAAGAGCGACGGGCGGCGCAAGCGCTCGGCT 473  
QY 139 eThrArgGlyArgGlnGlyArgHisHisValGlnGlyrProValGluPheProAlaTh 159  
Db 474 CGCTGCTGTCAACCGGCGGCGGAC-CGGCGGCTCGGCGATCGGCTACCTCCAGCAGTAC 532  
QY 159 rArgSerArgArgArgArgAlaSerSerSerAlaSerThrProTP-----SerSerPr 177  
Db 533 GCGGCGATCGGCTACCGCGGCGAAGGTCGCGCCAGTACGATGGGTGGGTGACACCC 592  
QY 177 oAlaSerValAspPheMetGluGluValThrProLySerValSerThrProSerValas 197  
Db 593 CGGG-----GCGTGGCGCGGAGTAA 613  
QY 197 pleuSerGlnAspGlyAspGlnGlnGlyMetAspThr--ThrGlnValAspAlaGlyse 216  
Db 614 CCGCTTCGAGTGTCTG-CACGCGCGGAGATGACGCGTACACGCGCACGACGACACCC 672  
QY 216 rArgAspGlyAspSerThrGluThrGlnAspAsp-----LySGluPheGlyIleGly-- 233

Db 673 GGACAGCCCGGCGGAGAGCGAGCTGTGACCGCTTACAGAGGTTGCCGAGGCGTGT 732  
QY 234 -----AspleuValTrpGlyLySileLySGlyPheSerTrpTrpProAlaMetVala 251  
Db 733 CCGGCGGAGCGG-----CGAAGCTGTGCGG 759  
QY 251 lSerTrpLysAlaThrSerLysArgGlnAlaMetProGlyMetArgTrpValGlnTrpPh 271  
Db 760 CCAGCTTCACCGGTGAGCGGCGACGACATGACGTCTTGGCGGCGTGT-----CT 813  
QY 271 eGlyAspGlyLysPheSerGluLysSerAlaAspLysLeuValAlaLeuGlyLeuPheSe 291  
Db 814 GGGCGACAGAGCTACCTACGTGGAGCGTCGTACGCGACCTTCTCGGCG----- 865  
QY 291 rGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrrArgLys----- 307  
Db 866 -----GCGACTACGCGGCTGTCTTCCCGACCGGACGCGGCGCTGCT 909  
QY 308 -----AlaMetTyrrHisThrLeuGluLysAlaArgValArgAlaGlyLysThr-- 323  
Db 910 CCTGAGCGGCGGATGACCCCTGCTGCGGCGGCGGCTGTAACCTGAGCAGACGGA 969  
QY 324 -----PheSerSerSerProGlyLysSerLeuGluAspGlnLeuLys--ProMetLeuG 341  
Db 970 GGGCTTCAGAGCGGCTTCCAGTCTTCCGGAAGAGCTGCTGTAAGCAGCAGCGA-CGTCC 1028  
QY 341 uTrpAlaHisGlyLysPheLysProThrGlyIleGlyLysProAsnLysLysG 361  
Db 1029 CCCTCGGCGGACAGGACACCAACCC-----CGACCAAGTCCGCAAGA 1070  
QY 361 nProValAlaAsnLysSerLysValArgArgSerAspSerArgAsnLeuGluProArgAr 381  
Db 1071 ACCTCAATC-----CTTCTTCAAGA-----CTTGACG 1100  
QY 381 rArgGluAsnLysSerArgArgArgThrThrAsnAspSerAlaAlaSerGlyuserProPr 401  
Db 1101 CGAAGCCCTGCGCGGCGGCGGCGCG--ACGGCGGCAAGCTCACCAGATCCCTGCGCAC 1159  
QY 401 oProLysArgLeuLysThrAsnSerTyrrGlyLysAspArgGlyGluAspGlyuse 421  
Db 1160 ACCGCGCGT-ATCGCGCGGATG-----TACGACGAGGCGGCTGGACAGCT 1206  
QY 421 rArgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAspArgCysLe 441  
Db 1207 GCGGAGTCCCTCACCTCGCGGATCAAGAGAACGCT----- 1246  
QY 441 uSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyLeuCySG 461  
Db 1247 -----GCGGCGCTGCTGATCCT 1263  
QY 461 nSerCysArgAspArgPheLeuGluLeuPheTyrrMetTyrrAspGluAspGlyTyrrGlnse 481  
Db 1264 CTCC-----GACAGCTACTACGAGCGGACGCGGACGCGC----- 1297  
QY 481 rTyrrCysThrValCysCysGluGlyArgGluLeuLeuLeuCysSerAsnThrSerCySG 501  
Db 1298 -----GGCTACAGCAACTGATGTTCGCCAACGCGCGCC----- 1330  
QY 501 sArgCysPheCysValGlyCysLeuGluValLeuValGlyAlaGlyThrAlaGluAspAl 521  
Db 1331 -----GTAACTGCTCGACCTC-----CCGCG 1353  
QY 521 aLysLeuGlnGluProTrpSerCyTyrrMetCysLeuPro-----GlnArgCysHisG 539  
Db 1354 CGCCTTCTCTCCCGGACGAGGTGCGGACGCGCTCCCGACTTCGAGAAAGCGGTCCCC 1413  
QY 539 yValLeuArgArgArgLysAspTrpAsnMetArgLeuGlnAspPhePheThrThrAspPr 559  
Db 1414 GGTCTTCGCGGAGGCGCTCGCTGTCTTCCCTGAACCTGGCGGTGACCGCGGTGAAGCC 1473  
QY 559 oAspleuGluGluPheGluProProLySerLeuTyrrProAlaAlaLysArgAr 579

```
Db 1474 C-----ACGGGGAGCGCCACCGC-----ATCGAGGCGCGCGGCCAC 1512
QY 579 gProlleArgValLeu-----Se 585
Db 1513 CCCGATCGTCTGTCGGCACACCGCGCCACCCCTACCGTGGGCGGAGGC 1572
QY 585 rLeuPheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeuGlyIleLysValG1 605
Db 1573 CCTCTCCGACGAGTCACCTCGGCGCACCTCTCTCACCTACGAG--GGAGACGCGCCACAC 1629
QY 605 uLysTyrIleAlaSerGluValCysAlaGluSer-----lleAlaValG1 620
Db 1630 CGCGTACGGCGCGCGGAGCTCTGTCATCGACTCCGCGATCAACACAGTACTGTGTGACCGG 1689
QY 620 YThrValLysHisGluGln 627
Db 1690 CACCGCCCGGAGGACGGCAAG 1711

RESULT 10
US-09-951-742-7
; Sequence 7, Application US/08951742
; Patent No. 6127144
; GENERAL INFORMATION:
; APPLICANT: Bartfeld, Daniel
; APPLICANT: Michael J. Butler
; APPLICANT: Dany Hadary
; APPLICANT: David Jeniah
; APPLICANT: Tim Krieger
; APPLICANT: Lawrence T. Malek
; APPLICANT: Gisela Soosmeyer
; APPLICANT: Eva Walczyk
; APPLICANT: Phyllis Krygsmann
; APPLICANT: Sheila Garven
; TITLE OF INVENTION: METHOD FOR EXPRESSION OF PROTEINS IN
; TITLE OF INVENTION: BACTERIAL HOST CELLS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/951,742
; FILING DATE: 16-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Bent, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 0189740/0140
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1821 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 104..1720
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 104..244
; FEATURE:
```

```
; NAME/KEY: mat_peptide
; LOCATION: 245..1720
US-08-951-742-7
Alignment Scores:
Pred. No.: 0.0134 Length: 1821
Score: 127.50 Matches: 145
Percent Similarity: 33.95% Conservative: 75
Best Local Similarity: 22.38% Mismatches: 251
Query Match: 2.78% Indels: 182
DB: 3 Gaps: 30

US-09-720-086-6 (1-859) x US-08-951-742-7 (1-1821)
QY 37 AlaProSerProProValLeuGluAlaIleCysThrGluProValCysThrPro-GluTh 56
Db 135 GCACCGCGCGCACCCGTTTCCGGGCCACAGCTGCTCACCGCGCGCTGCTCGCCACCGCT 194
QY 56 rArgGlyArgSerSerArgLeuSerLysArgGluValSerSerLeuLeuAsnTy 76
Db 195 GCTCGGCGG-----GGGCGGCTCGAGCTCGCGCGGATCCCCCGCGCGCAAGGGCGG 245
QY 76 rThrGlnAspMetThrGlyAspGlyAspArgAspAspGluValAspAspGlyAsnGlyse 96
Db 246 CCGGCGCGCGAGGCGGCGCACCGCGACCTGACCCCTGCCGAA-----GGCCACGC 299
QY 96 rAspIleLeuMetProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSe 116
Db 300 CCGCGGAGCTGCTCCCGCTACTACGAGCAGAGGCTCGGCTGGCGGCGACTGCGG-----CG 353
QY 116 rProAlaValArgThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnArg----- 134
Db 354 TCCCGGGCTTCCAGTGCAGCCACCATGAAGCCCGCTCGACTACGCCAAGCCCGCGGAGC 413
QY 135 -----AlaSerProArg11 139
Db 414 GCGAGCTCGCGGCTCGCGGTGGCCCGCAAGAGCCACCGCGCGCGCGCGCTCGCT 473
QY 139 eThrArgGlyArgGlnGlyArgHisValGlnGluTyrProValGluPheProAlaTh 159
Db 474 CGCTCTGCTCAACCGCGGCGGAC-CGGCGGCTCGCGCATCGGCTACCTCCAGCGAGTAC 532
QY 159 rArgSerArgArgArgAlaSerSerSerAlaSerThrProTrp-----SerSerPr 177
Db 533 GCGGCGATCGGCTACCGCGGAGAGTCCGCGCCAGTACGACATGGTGGCGGTGACCC 592
QY 177 oAlaSerValAspPheMetGluGluValThrProLysSerValSerThrProSerValAs 197
Db 593 CCGG-----CGTGGCGCGCGAGTGAA 613
QY 197 pLeuSerGlnAspGlyAspGlnGluGlyMetAspThr---ThrGlnValAspAlaGluSe 216
Db 614 CCGGTCGAGTGGCTG-GACGGGCGCGAGATGGACGCTACACGCGCACCGACGTCAACCCC 672
QY 216 rArgAspGlyAspSerThrGluTyrGlnAspAsp-----LysGluPheGlyIleGly-- 233
Db 673 GACACGCGCGGCGGAGACGAGCGAGTGGTTCGACGCGCTACAGAGAGTTCGCGAGGGGTG 732
QY 234 -----AspLeuValTrpGlyLysIleLysGlyPheSerTrpTrpProAlaMetValVa 251
Db 733 CCGGCGCGGACCGG-----CCGAGCTGCTGCG 759
QY 251 lSerTrpLysAlaThrSerLysArgGlnAlaMetProGlyMetArgTrpValGlnTrpPh 271
Db 760 CCACGTCCTCCAGCGTGGGCGGCGCACGAGATGGAGCTCTGCGCGCGGTG-----CT 813
QY 271 eGlyAspGlyLysPheSerGluIleSerAlaAspLysLeuValAlaLeuGlyLeuPheSe 291
Db 814 GGGCGACGAGAGAGTACCTACGTGGGAGCGTACGCGCACCTTCCTGCGG----- 865
QY 291 rGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLys----- 307
Db 866 -----GCGACCTACGCGGCTGTCTTCCCGACCGGCGGCGCGCTGCTGT 909
```

```

QY 308 -----AlaMetTyrHisThrLeuGluValAlaArgValArgAlaGlyIysThr-- 323
Db 910 CTTGAGACGGCGGATGAGCCCTCGTCCGCCGCCCTGAACTTGAGAGACAGCGGA 969
QY 324 -----PheSerSerSerProGlyIleSerLeuGluAspGlnLeuLys---ProMetLeuG1 341
Db 970 GGGCTTCGAGACGGCGGCTTCAGTCTTCGCGAAGACTGCGTGAAGACAGCGGA-CTGCC 1028
QY 341 uTTPAlaHisGlyGlyPheLysProThrGlyIleGluGlyLeuLysProAsnLysLysG1 361
Db 1029 CCCTCGGCGACAGACAGACACCC-----CGACCGAGCTGGCGCAGA 1070
QY 361 nProValValAsnLysSerLysValArgArgSerAspSerArgAsnLeuGluProAlaArg 381
Db 1071 ACCCTCAAGTC-----CTTCTTGACGA-----CCTGACAG 1100
QY 81 gArgGluAsnLysSerArgArgThrThrAsnAspSerAlaAlaSerGluSerProPr 401
Db 1101 CGAAGCCCTCGCCCGCGCGCGAGCGCG-ACGGCCGCAAGCTCACCGAATCCTCGCCACC 1159
QY 401 oProLysArgLeuLysThrAsnSerTyrGlyLysAspArgGlyGluAspGluG1Se 421
Db 1160 ACCGCGCGT-ATCGCCGCGATG-----TACGACGAGGCGCGCTGGCAGCAGCT 1206
QY 421 rArgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAspArgCysLe 441
Db 1207 GCGGAGATCCCTACCTCGCGCATGAGAGAGAGACGCT----- 1246
QY 441 uSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyLysLeuG1 461
Db 1247 -----GGGGGCTGCTGATCTT 1263
QY 461 uSerCysArgAspArgPheLeuGluLeuPheTyrMetTyrAspGluAspGlyTyrGlnSe 481
Db 1264 CTCC-----GACAGCTACTACGAGCGCGAGCGCGACGCGC----- 1297
QY 481 rTyrCysThrValCysCysGluGlyArgGluLeuLeuLysSerAsnThrSerCysCy 501
Db 1298 -----GGCTACGACGAACCTGATGTTGCCACGCGCGC----- 1330
QY 501 sArgCysPheCysValGluCysLeuGluValLeuValGlyAlaGlyThrAlaGluAspAl 521
Db 1331 -----GTGAACCTGCTCGACCTC-----CCCGC 1353
QY 21 aLysLeuGluGluProTyrSerCysTyrMetCysLeuPro-----GlnArgCysHisG1 539
Db 1354 GCGCTTCTCTCCCGGAGCGAGTGGCGAGCGCTCCCGACTTCGAGAGAGCGTCCCC 1413
QY 539 yValLeuArgArgArgLysAspTyrAsnMetArgLeuGlnAspPhePheThrThrAspPr 559
Db 1414 GGTCTTCGCGAGGCGCTCGCTGCTCTCCCTGAACTGCGCGTACTGCGCGGGAAGCC 1473
QY 559 oAspLeuGluGluPheGluProProLysLeuTyrProAlaIleProAlaAlaLysArgAr 579
Db 1474 C-----ACGGGGAGAGCGGCACCGC-----ATCGAGGCGGCGCGCGCAC 1512
QY 579 gProLysArgValLeu-----Se 585
Db 1513 CCCGATCGCTGCTGCGGACACCGCGACCGCGCACCCCTTACCGCTGGGCGGAGGC 1572
QY 585 rLeuPheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeuGlyIleLysValG1 605
Db 1573 CCTTCGCGACAGCTACCTCGGCGACCTCTCACTACAG--GGAGACGGCGCACAC 1629
QY 605 uLysTyrIleAlaSerGluValCysAlaGluSer-----IleAlaValG1 620
Db 1630 CGCGTAAGCGCGGCGAGCTCTGCATCGACTCGCGCATCAACAGCTACCTGCTGACCGG 1689
QY 620 yThrValLysHisGluGlyGln 627
Db 1690 CACCGCGCGGAGAGCGCAAG 1711

```

```

RESULT 11
US-08-973-675-1
; Sequence 1, Application US/08973675
; Patent No. 5985283
; GENERAL INFORMATION:
; APPLICANT: HATEBOER, GUIS
; APPLICANT: BERNARDS, RENB
; TITLE OF INVENTION: ADENOVIRUS E1A-ASSOCIATED PROTEIN BS69,
; TITLE OF INVENTION: INHIBITOR OF E1A-TRANSACTIVATION
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHAYE P.C.
; STREET: 1100 NORTH GLEBE ROAD
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/973,675
; FILING DATE: 12-DEC-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WILSON, MARY J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 620-31
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4100
; TELEFAX: (703) 816-4000
; INFORMATION FOR SEO ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2623 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 150..1835
; US-08-973-675-1
;
Alignment Scores:
Pred. No.: 0.0278 Length: 2623
Score: 127.00 Matches: 67
Percent Similarity: 30.42% Conservative: 41
Best Local Similarity: 18.87% Mismatches: 113
Query Match: 2.77% Indels: 134
Gaps: 12
US-09-720-086-6 (1-859) x US-08-973-675-1 (1-2623)
QY 234 AspLeuValITripGlyLysIleLysGlyPheSerTrrPrrProAlaMetValAlaSerTrp 253
Db 873 GAGCTGGTTGGCTAAATGAATGAAGTTTGGGTTTGGCGCAGCAAGTCATGCAG--- 929
QY 254 LysAlaThrSerLysArgGlnAlaMetProGlyMetArgTrpValGlnTrpPheGlyAsp 273
Db 929----- 929
QY 274 GlyLysPheSerGluIleSerAlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHis 293
Db 930-----AAAGAAGACAATCAAGTCGACGTCCTTCTTGGCACACAC 971
QY 294 PheAsnLeuAla-----ThrPheAsn----- 300
Db 972 CACCAAGAGGCTGATCTCTTGAAACAATTCAGATATCAAGTCAACATTCATCGG 1031
QY 301 -----LysLeuValSerTyrArgLysAlaMetTyrHis---ThrLeuGluLys 315

```

Db 1032 CTGACGCTGACGCGAGTATGGTTGGAAAAAGCCCTGTGATGAGCTGGAGCTGCATCAG 1091  
Qy 316 AlaArgValArgAlaGlyLys----- 322  
Db 1092 CGTTTCTTACGAGAAGGAGATTTTGGAAATCTTAAGAAATGAGGACCGAGGTGAGGAAGAG 1151  
Qy 323 -----ThrPheSerSerProGlyGluSerLeuGluAspGlnLeuLysProMet 339  
Db 1152 GCAGAAATCCAGTATCTCTCCACAGTATGACAGTAAAGTCACTCAAGAACCAAGA 1211  
Qy 340 LeuGluTrpAlaHisGlyGlyPheLysProThrGlyLeuGlyLeuLysProAsnLys 359  
Db 1212 GCAAAAG-----AAAGGACGACGTAATCAAAAGTGTGGAGCCCAAAAG 1253  
Qy 360 LysGlnProValValAsnLysSerLysValArgSerAspSerArgAsnLeuGluPro 379  
Db 1254 GAAGAACGACGACCTGAAACAGACAGTAAAGTTCTAGCCAGGAATACCCAGCATGCT 1313  
Qy 380 ArgArgArgGluAsnLysSer----- 386  
Db 1314 CAGCCCATCGAAAAAGTCTCCGTGTCACTCAGACAAAGAAAGTTAAGTGCCTCTTACCA 1373  
Qy 387 -----ArgArgArgThrThrAsnAspSerAlaAlaSerGluSerProPro 401  
Db 1374 AGAATGCTGCATCGGACGACCCAGACCAACAGGCGGTGTGTCAGAGCATGTGCCAT 1433  
Qy 402 ProLysArgLeuLysThr-----AsnSerTyrGlyLysAspArgGlyGlu----- 417  
Db 1434 GACAAATACACCAAGATCTTCAATGACTTC-----AAAGCCGGATGAAGTCGGACAC 1487  
Qy 418 -----Asp 418  
Db 1488 AAGCGGAGACAGACGCTGTCTCCGAGAAGCTCTCGAGAAGCTGCGTCTCGAAATGGAA 1547  
Qy 419 GluGluSerArgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAsp 438  
Db 1548 GAAGAAAGACAAAGCTGTAAATAAAGCTGTAGCCCAACATGCGGGTGAGATGGAGAGA 1607  
Qy 439 ArgCysLeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGly 458  
Db 1608 AAATGTAGCAAGTAAGGAAAAAG----- 1631  
Qy 459 LeuCysGlnSerCysArgAspArgPheLeuGluLeuPhe----- 471  
Db 1632 -----TGTAAGGAGGAATTTGTAGAAGAAATCAAGAAGCTGGCAACACACAC 1679  
Qy 472 ---TyrMetTyrAspGluAspGlyTyrGlnSerTyrCysThrValCysCysGluGlyArg 490  
Db 1680 AAGCAACTGATTCTTCAGACCAAGAAAGACAGTGTGTCTACAACTGTGTAGGAGGAGGCC 1739  
Qy 491 GluLeuLeuLeuCysSerAsnThrSerCysCysArgCysPheCys 505  
Db 1740 ATGTACCCTGCTGTGGAACACATCTCTACTGCTCCATCAAGTGC 1784

## RESULT 12

US-09-453-702B-96/c  
; Sequence 96, Application US/09453702B  
; Patent No. 6365723

## GENERAL INFORMATION:

; APPLICANT: Blattner, Frederick R.  
; Burland, Valerie  
; Plunkett, Guy  
; Welch, Rod

TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157

NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESS: Quares & Brady

STREET: 1 South Pinckney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53701-2113

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 8.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/453,702B  
FILING DATE: 03-Dec-1999  
CLASSIFICATION: <Unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/110,955

FILING DATE: 04-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27386

REFERENCE/DOCKET NUMBER: 960296.95017

TELECOMMUNICATION INFORMATION:

TELEPHONE: (608) 251-5000

TELEFAX: (608) 251-9166

INFORMATION FOR SEQ ID NO: 96:

SEQUENCE CHARACTERISTICS:

LENGTH: 34063

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 96:

US-09-453-702B-96

## Alignment Scores:

Pred. No.:	2,99	Length:	34063
Score:	125.50	Matches:	58
Percent Similarity:	40.69%	Conservative:	36
Best Local Similarity:	28.11%	Mismatches:	88
Query Match:	2.73%	Indels:	49
DB:	4	Gaps:	11

US-09-720-086-6 (1-859) x US-09-453-702B-96 (1-34063)

Qy 573 IleProAlaAlaLysArgArg-----ProIleArgValLeuSerLeuPheAspGlyIle 590

Db 31583 ATAGAACTAATAAGACGCTTGCACATATATGAATCAATATAGATTGTTTCTGCGCTT 31524

Qy 591 AlaThrGlyTyrLeuValLeuLysGluLeuGlyLleLysValGluLysTyrIleLaser 610

Db 31523 GGAGGCTTAAGTCTTGGTCTGCACGCTGCGATTGATGTTAAA-----ATGCGAGTT 31470

Qy 611 GluValCysAlaGluSerIleAlaValGlyThrValLysHisGluGlyGlnIleLysTyr 630

Db 31469 GAAATTGATCAACATGCTATTAACTACACCAATTAATTTCCAAAGAAAGTTTGCACGTC 31410

Qy 631 ValAsnAspVal-----ArgLysIleThrLysLysAsnIleGluGluTrpGly 646

Db 31409 CAAGAAGATGTTCTTTACTAATATGAGAAATTAATTAAGGTTTATTAATAACGATATG 31350

Qy 647 ProPheAspLeuValIleGlySerProCysAsnAspLeuSerAsnVal----- 663

Db 31349 CCCATAGATGGTATTATTGGCGGCTCTCCGTCGCAAGGATTAGTTCAATAGCAAGGG 31290

Qy 664 AsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPhePheGluPheTyrHis 683

Db 31289 AATCCTGAT-----GATAGCAGGAATCAGCTTTACATGCACTTTCTACCGT 31245

Qy 684 LeuLeuAsnTyrThrArgProLysGluGlyAspAsnArgProphePheTrpMetPheGlu 703

Db 31244 TTAGTATCA-----GAATTACAGCCATTATTCTTTTGGCAGAA 31206

Qy 704 AsnValValAlaMetLysValAsnAspLysLysAspLysArgPheLeuAlaCysAsn 723

Db 31205 AATGTTCCAGGTATT---ATGCAAGAGAAATATTTCTGCATTACAAATAAAGCATTTTAT 31149

Qy 724 ProVal-----MetIleAspAlaLysValSerAlaAlaHis----- 736

Db 31148 TTGTTAGCGGTGATTATGATATCTTGATCCCATCAAGTAAAGCATCTGATTATGCT 31089  
QY 737 -----ArgAlaArgTyrPheTyrPGLyAsnLeuProGlyMetAsnArgProVal 752  
Db 31088 GCTCCAACTATTGCAACTAGATATTTT-----ATC 31056  
QY 753 MetAlaSerLysAsnAspLysLeuGlnAspCysLeuGlnPheSerArgThrAla 772  
Db 31055 GGTGTAAAAAATCATTTGAAGCTTGATATTTCAGAT-----GAAGTATT 31011  
QY 773 LysLeuLysLysValGlnThrIleThrThrLys 783  
Db 31010 ATGCCTAAATGATTGATCCGTTACTGTAAAA 30978  
RESULT 13  
US-09-103-840A-1  
Sequence 1, Application US/09103840A  
Accession No. 6294328  
LOCAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103,840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37Rv  
US-09-103-840A-1  
Alignment Scores:  
Pred. No.: 1.11e+04 Length: 4411529  
Score: 125.50 Matches: 156  
Percent Similarity: 30.39% Conservative: 61  
Best Local Similarity: 21.85% Mismatches: 245  
Query Match: 2.73% Indels: 258  
DB: 4 Gaps: 36  
US 20-086-6 (1-859) x US-09-103-840A-1 (1-4411529)  
QY 60 ArgSerSerSerArgLeuSerLysArgGluValSerSerLeuAsnTyrThrGlnAsp 79  
Db 1605907 CGACGACGCAATAGCTGCTACCGGACGACGATTCGCGAGTTCGTGCTGAACCGCAG 1605966  
QY 80 MetThrGlyAspGlyAspArgAspArgGluVal-----Asp 91  
Db 1605967 CTACGTGCGCACCGCGCGCGGATCGATCTGTCGACCCACCAACCATTCATACCGGATGAT 1606026  
QY 92 AspGlyAsnGlySerAspIleLeuMetProLysLeuThrArgGluThrLysAspThrArg 111  
Db 1606027 CCGAGCGCATGAACT----- 1606041  
QY 112 ThrArgSerGluSerProAlaValArgThrArgHisSer----- 124  
Db 1606042 ATTCGCGACAGATCTCGACGATCCCGATGCCGCTTCGGGTGACGATGCGCTGAGGT 1606101  
QY 125 -----AsnGlyThrSerSerLeuGlnArgGlnArgAlaSerProArgIleThrArg 141  
Db 1606102 CTGCCGCTGATGAGGACCGCGCGGTGCTCCGCAACAAATAGTCTCGGTAAACGCCCG 1606161  
QY 142 GlyArgGlnGlyArgHisValGlnGluTyrProValGluPheProAlaThrArgSer 161  
Db 1606162 GTCGCTTCATGCGTGGCGAC-----AATCGTGT 1606191  
QY 162 ArgArgArgArgAlaSerSerSerAlaSerThrProThrSerSerProAlaSerValAsp 181

Db 1606192 CGTAGCTTCGCTGTCGCGCATATGCTCTCGACGATTCGCGCCT----- 1606239  
QY 182 PheMetGluGluValThrProLysSerValSerThrProSerValAspLeuSerGlnAsp 201  
Db 1606240 -----AACGTTTGCCACGCTCAACAAACCCGACAGAA-----AGCCAGCCG 1606281  
QY 202 GlyAspGlnGlyMetAspThrThrGlnValAspAlaGluSerArgAspGlyAspSer 221  
Db 1606282 GTCGCGCAGAGGGGGG-CGAGCCCGGCGTATACCAATTCGGTGCCTC-GGTTCTAGT 1606339  
QY 222 ThrGluTyrGlnAspAspLysGluPheGlyIleGlyAspLeuValTyrGlyLysIle--- 240  
Db 1606340 -----TGCGCGCTATCATCGTTGCCAGCGGGTGTGCGGA 1606375  
QY 241 -----LysGlyPheSer-----TyrThrProAlaMetValValSer 252  
Db 1606376 AGGTGCGTATGCTGCTGTTTCGCGGTGCCAGAGATGTGGCGCAACCGCTTCGAT 1606435  
QY 253 TyrLysAlaThrSerLysArgGlnAlaMetProGlyMetArgTyrValGln----- 269  
Db 1606436 TGGCAGACCTCGAGCGCGCTGAGCGAGGCCACCGCGCGCGCTATCCCGACACAC 1606495  
QY 270 -----TyrPheGlyAspGlyLysPheSerGluIleSerAlaAspLysLeuValAlaLeu 287  
Db 1606496 AAGTACTGACC-GCGGCGCC-----GATGAGTGTGCGCGCATCGCGAGTTGTT 1606548  
QY 288 GlyLeuPheSerGlnHisPheAsnLeuAlaThr----- 298  
Db 1606549 GGTGCGCAGCGCCAAAGATTTCAGCGCTCAGCGCCAGCATCGCGCTTCATGACCG 1606608  
QY 299 PheAsnLysLeuValSerTyrArgLysAlaMetTyrHisThrLeuGluLysAlaArgVal 318  
Db 1606609 TTGCTGCGGCGCTTAAGCCCGCAGCGGCTGTGTATGTGACCCCGAGCGCCAAACGC- 1606667  
QY 319 ArgAlaGlyLysThrPheSerSerSerProGlyGluSerLeuGluAspGlnLeuLysPro 338  
Db 1606668 CCGCGCTGCT-----GGACACCGCGCGCCACCGCGCGCTCGAGATT 1606706  
QY 339 MetLeuGluTyrAlaHisGlyGlyPheLysProThrGlyIleGlu----- 353  
Db 1606707 GGGGTCAGTGGGCGCCACCGCGCT-----GATTCGTGCGCTCCACCGCAACCCCGGAC 1606760  
QY 354 -----GlyLeuLysProAsnLysLysGlnProValAlaAsn--- 365  
Db 1606761 GCCCTTGACATACATGACACGAGGTCTACGACCGCTACATCGCACCCCACTACTGGGCTA 1606820  
QY 366 -----LysSerLysValArgArgSerAspSerArg 375  
Db 1606821 TGGCTTTCCGCGCTGTACACCGCCCGCAGTTTCAGCCGTGACCGGCATCCCGACCT 1606880  
QY 376 AsnLeuGluPro-----ArgArgArgGluAsnLysSerArgArgThrThrAsnAsp 393  
Db 1606881 GACCTACGACCAATCGGTGCGCCAGAGCGCGCGCTATCTTCACACCGC-----GAT 1606931  
QY 394 SerAlaAlaSerGluSerProPro----- 401  
Db 1606932 CATGACAGCAAGTCGCGCGCGCAATGACGTTGTGTGTTGGGTTTCTCGACAGGCGCGCT 1606991  
QY 402 -----ProLysArgLeuLysThrAsnSerTyrGlyLysAspArgGlyGluAsp 418  
Db 1606992 GGTGCGCACCTTGAAATGCGCCATCTGCGCAAGCCTGCGCGCGCGCGCGCGAGTCC 1607051  
QY 419 GluLysSerArgGluArgMetAlaSerGluValThrAsnLysGlyAsnLeuGluAsp 438  
Db 1607052 GAATCACTCTGTTGTGTAATGTGGGCAACCCCAACCAACCG----- 1607099  
QY 439 ArgCysLeuSerCysGlyLysLysAsnProValSerPheHis-----ProLeuPheGluGly 457  
Db 1607100 -----GGGCACTCTGCGCGGTTTCCGGTCTGTACTGCAGTGCGCTCGG 1607144  
QY 458 GlyLeuCysGlnSerCys-----ArgAspArgPheLeuGluLeuPhe 471  
Db 1607145 CTTGACGTTCAACGGTGCACACCCCGGACACCGACGACTACGACGAC----- 1607189





```

Db 2827 -----GACGGGTTCTCG----- 2838
Qy 258 LysArgGlnAlaMetProGlyMetArgTrpValGlnTrpPheGlyAspGlyLysPheSer 277
Db 2839 -----TGGGGCATGCACTGG-----GATGGGCACTTCCCG 2868
Qy 278 GluIleSerAlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAla 297
Db 2869 -----GGC 2871
Qy 298 ThrPheAsnLysLeuValSerTrpArgLysAlaMetTrpHisThrLeuGlnLysAlaArg 317
Db 2872 TCCTTCATCGCTGCGCGGCTTGGCCCTCAA-----CATCCGAAGCAT 2916
Qy 318 ValArgAlaGlyLysThrPheSerSerSerProGlyLysLeuGlnAspGlnLys 337
Db 17 GTTACCATCGGCTCCAG-----GACATGATGCAC 2946
Qy 338 ProMetLeuGlnTrpAlaHisGlyLysPhe-----LysProThrGlyIleGlnGlyLeu 355
Db 2947 ACCCCGAGAGATGGAATACAGTGGAGTTGGGTGGAGTCCGACTCATGG----- 2994
Qy 356 LysProAsnLysLysGlnProValAlaAsnLysSerLysValArgArgSerAspSerArg 375
Db 2995 -----TCGTTGGCTCGCTCAGTGAAGTGC-----AACGA 3027
Qy 376 AsnLeuGlnProArgArgArgGlu----- 383
Db 3028 GAGCAGACCCCTCGCGGAGAGATGGCGGCACACGTCACCCCAATACGCCAG 3087
Qy 384 -----AsnLysSerArgArgArgThrThrAsnAspSerAlaAlaSerGlnSerProPro 401
Db 3088 CTGTTGCCCAAGACATGACACTATTAACAATACGCTCACTACCTCCCTAATACGCCG 3147
Qy 402 ProLysArgLeuLysThrAsn-----SerTrpGlyGlyLysAsp 414
Db 3148 CCCGATCCGCTCGAGCAGCAGCATTCGTCGCCCGCCGCGGAGTCCGGGGAGAGCAG 3207
Qy 415 ArgGlyGlnAspGlnGlnSerArgGlnArgMetAlaSerGlnValThrAsnAsnLysGly 434
Db 3208 AACGGCGAC-----CAAGCGACG 3225
Qy 435 AsnLeuGlnAspArgCysLeuSerCys-----GlyLysLysAsn 447
Db 4326 AACGACCCGACCACTGACCACTCTCACTCAAAACCACTCCGCTGGCGCTCGGAC 3285
Qy 448 ProValSerPheHisProLeuPheGlnGly---GlyLeuCysGlnSerCysArgAspArg 466
Db 3286 CCA-----GAGGGCGACGCACTGTGCATGCCCTGC---GGTTG 3321
Qy 467 PheLeuGlnLeu 470
Db 3322 TTTTGAAATTG 3333

RESULT 15
US-09-079-415-5
; Sequence 5, Application US/09079415
; Patent No. 6013452
; GENERAL INFORMATION:
; APPLICANT: Christensen, Tove
; APPLICANT: Lembeck, Jan
; TITLE OF INVENTION: A Fungus Wherein The area, pepc and/or
; TITLE OF INVENTION: pepc Genes Have Been Inactivated
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 60134520 No. 6013452disk of No. 6013452th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/079,415
; FILING DATE: 14-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 4657,204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5643 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus oryzae
; STRAIN: IFO4177
; FEATURE:
; NAME/KEY: Intron
; LOCATION: 2701..2769
; FEATURE:
; NAME/KEY: CDS
; LOCATION: John(2282..2700, 2770..4949)
; US-09-079-415-5

Alignment Scores:
Pred. No.: 0.16 Length: 5643
Score: 125.00 Matches: 101
Percent Similarity: 30.56% Conservative: 53
Best Local Similarity: 20.04% Mismatches: 142
Query Match: 2.72% Indels: 208
DB: 3 Gaps: 24

US-09-720-086-6 (1-859) x US-09-079-415-5 (1-5643)
Qy 9 AsnGlnGlnGlnGlyLysArgGlyTrpGlnGlnCysIleIleValAsnGlyAsnPheSer 28
Db 3419 AGCGAGTCTCCGATGGCTCCGATCATCG-----TTTGGC 3454
Qy 29 AspGlnSerSerAspThrLysAspAlaProSerProProValLeuGlnAlaIleCysThr 48
Db 3455 AACCTCTATTCGATAC-----CCGGTGGCTTCGTCCTCACTCG 3496
Qy 49 GluProValCysThrProGlnThrArgGlyArgArgSerSerSerArgLysSerArg 68
Db 3497 ACGGATTTCTTCTCTCCACCGCATGAGCTACAGTCCACGGCATCC----- 3544
Qy 69 GluValSerSerLeuLeuAsnTrpThrGlnAspMetThrGlyAspGlyAspArgAspAsp 88
Db 3545 -----ACGCCGACCCCACTACGACGAGGAGGACCATTCGCTT 3580
Qy 89 GluValAspAspGlyAsnGlySerAspIleLeuMetProLysLeuThrArgGlnThrLys 108
Db 3581 TATTTGATATGCGCTGGG----- 3601
Qy 109 AspThrArgThrArgSerGlnSerProAlaValArgThrArgHisSerAsnGlyThrSer 128
Db 3602 GACGGCGGACCCAGCGCCGATTCGAACTATTTGCGCATCGGTCCAACTGTCTGCT 3661
Qy 129 SerLeuGlnArgGlnArgAlaSerProArgIleThrArgGlyArgGlnGlnArgHisHis 148
Db 3662 TCGCTGCAG-----CCTCG----- 3676

```



Qy 149 ValGlnGluTyrProValGluPheProAlaThrArgSerArgArgArgArgAlaSerSer 168  
Db 3677 -----TATATGTTCAACCAAGAACACCATCAACAGGCGAGTCG 3715  
Qy 169 SerAla-----SerThrProTrpSerSerProAlaSerValAlaPheMetGlu 184  
Db 3716 TCGACGGTGTCATTCGCCGAGCTACCCCAATTCGCCAGCCGCAACATGTGGACCCCACTCAG 3775  
Qy 185 GluValThrProLysSerValSerThrProSer-----ValAsp 197  
Db 3776 GTGTTGAACGCCACCAATTACTCGACCGGCAACTCCACCATTACCGGCGCCATGTTTCA 3835  
Qy 198 LeuSerGlnAspGlyAspGlnGluGlyMetAspThrThrGlnValAlaGluSerArg 217  
Db 3836 TTTGGAGCCGATTCAGATAACGAGGATGACGATGTCATCAGCTCCGAGCGGCTGGT 3895  
Qy 218 AspGlyAspSerThrGluTyrGlnAspAspLysGluPheGlyIleGlyAspLeuValTrp 237  
Db 3896 CTGGCGATGCCGACTGAATATGGGACGAG----- 3925  
Qy 238 GlyLysIleLysGlyPheSerTrpTrpProAlaMetValSerTrpLysAlaThrSer 257  
Db 3926 -----GACGGTTCTCG----- 3937  
Qy 258 LysArgGlnAlaMetProGlyMetArgTrpValGlnTrpPheGlyAspGlyLysPheSer 277  
Db 3938 -----TCGGGCATGCAGTGG-----GATGGCAGTTCCCG 3967  
Qy 278 GluIleSerAlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAla 297  
Db 3968 -----GGC 3970  
Qy 298 ThrPheAsnLysLeuValSerTyrArgLysAlaMetTyrHisThrLeuGluLysAlaArg 317  
Db 3971 TCCTTCCATTCCTGCCGGCTTTGCCCTCAA-----CATCGCAAGCAT 4015  
Qy 318 ValArgAlaGlyLysThrPheSerSerProGlyGluSerLeuGluAspGlnLeuLys 337  
Db 4016 GTTACCATCGGGTCCACG-----GACATGATGGAC 4045  
Qy 338 ProMetLeuGluTrpAlaHisGlyGlyPhe-----LysProThrGlyIleGluGlyLeu 355  
Db 4046 ACCCCCGAGGAGTGAATCAGCGTGGCAGTTTGGTTCGACTCATGGG----- 4093  
Qy 356 LysProAsnLysLysGlnProValValAsnLysSerLysValArgArgSerAspSerArg 375  
Db 4094 -----TCGGTGGCTTCGTCAGTGCAGTGGC-----AACCGA 4126  
Qy 376 AsnLeuGluProArgArgArgGlu----- 383  
Db 4127 GAGCAGGACCTTCGCGCAGAGAGATTGCCCGCACACGTCCACCCCAATACGCCCCAG 4186  
Qy 384 -----AsnLysSerArgArgArgThrThrAsnAspSerAlaAlaSerGluSerProPro 401  
Db 4187 CTGTTCCGCCCAAGCATGCTCTAATAACAATACGTCTCATACCTCCCTAATACGCCG 4246  
Qy 402 ProLysArgLeuLysThrAsn-----SerTyrGlyLysAsp 414  
Db 4247 CCCGAGTCCGCCCTGAGCAGCGAGTTCGTCGCCCGCGGCGAGTCCCGGGGCGAGCAAG 4306  
Qy 415 ArgGlyGluAspGluGluSerArgGluArgMetAlaSerGluValThrAsnAsnLysGly 434  
Db 4307 AACGGCGAC-----CAAGGCAGC 4324  
Qy 435 AsnLeuGluAspArgCysLeuSerCys-----GlyLysLysAsn 447  
Db 4325 AACGGACCGACCACTGCACCACTGCTTCACTCAAAACCACTCCGCTGTGGCGTCGGAAC 4384  
Qy 448 ProValSerPheHisProLeuPheGluGly-----GlyLeuCysGlnSerCysArgAspArg 466  
Db 4385 CCA-----GAGGCGCAGCAGTGTGCAATGCTCTGC-----GGGTG 4420  
Qy 467 PheLeuGluLeu 470

Db 4421 TTTTGAATTG 4432

Search completed: November 25, 2002, 02:56:29  
Job time : 4354.34 secs

Sequence 23	App
Sequence 23	App
Sequence 22	App
Sequence 22	App
Sequence 22	App
Sequence 22	App
Sequence 5	App1
Sequence 5	App1
Sequence 5	App1
Sequence 5	App1
Sequence 1	App1
Sequence 93	App
Sequence 93	App
Sequence 3	App1
Sequence 3	App1
Sequence 2	App1
Sequence 22	App1
Sequence 7	App1
Sequence 66	App
Sequence 1	App1
Sequence 74	App1
Sequence 70	App
Sequence 68	App1
Sequence 64	App1
Sequence 76	App1
Sequence 72	App1
Sequence 30	App1
Sequence 5	App1
Sequence 19	App1
Sequence 12	App1
Sequence 12	App1
Sequence 4	App1
Sequence 1	App1

## ALIGNMENTS

RESULT 1  
 Sequence 47, Application US/09276531  
 Patent No. 6183968  
 GENERAL INFORMATION:  
 APPLICANT: Bandman, Olga  
 APPLICANT: Lal, Preeti  
 APPLICANT: Hiltman, Jennifer L.  
 APPLICANT: Yue, Henry  
 APPLICANT: Reddy, Roopa  
 APPLICANT: Guegler, Karl J.  
 APPLICANT: Baughn, Mariah R.  
 TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
 NUMBER OF INVENTION: RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION  
 NUMBER OF SEQUENCES: 134  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
 STREET: 3174 PORTER DRIVE  
 CITY: PALO ALTO  
 STATE: CALIFORNIA  
 COUNTRY: USA  
 ZIP: 94304  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/276,531  
 FILING DATE: Herewith  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/079,677  
 FILING DATE: March 27, 1996  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:

```

? NAME: Lynn E. Murry, Ph.D.
? REGISTRATION NUMBER: 42,918
? REFERENCE/DOCKET NUMBER: PA-0008 US
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (650) 855-0555
? TELEFAX: (650) 845-4166
? INFORMATION FOR SEQ ID NO: 47:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2077 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? IMMEDIATE SOURCE:
? LIBRARY: TESTU702
? CLONE: 1271435
?
US-09-0276-531-47

```

Alignment Scores:

```

File: NO.1: 2077
Score: 748.00
Percent Similarity: 81.95%
Best Local Similarity: 73.65%
Query Match: 15.16%
DB: 4
Gaps: 1

```

US-09-120-086-7 (1-912) X US-09-276-531-47 (1-2077)

Db	3	GGCGGCGNANTGCACGATCTNNAAAT-GTGAATCA-GCCAGGAAAGGCTGTATGACGAC	60
QY	727	ThrgIyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgProLysGluGly	726
Db	61	AGG---CGGCTCTTCTTCGAATTTC-CACCTGCTGAATTACTACGCCCCCAAGAGGCT	116
QY	747	AspAspArgProPhePhePheTyrLeuPhePheGluAsnValValAlaMetClyValSerAspLys	766
Db	117	GATACCCGCGCGCTTCTTCTTGAGATGTTTGAGAAATGTTGAGNCTCGAGCCGATTCGGCGCA	176
QY	767	ArgAspLysSerArgPheLeuGluSerAsnProValMetLysAlaLysGluValSer	786
Db	177	GGGAGACATCTACGGGTCTCTGAGGTATATCCAGTGATGATGATGCCATCAAAAGTTTCT	236
QY	787	AlaAlaHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMetAsnArgProLeuAla	806
Db	237	GCTGCTCACAGGCGCCGATACTTCTTGCGGCAACTACCCGGGATGAACAGCCCGTGATA	296
QY	807	SerThrValAsnAspLysLeuGluLeuGlnGlyLeuGluHisGlyValArgIleAlaLys	826
Db	297	GCATCAAGAATGATATTAACCTCGNGCTGCGAGGACTCTTGGAATCAATAGATAGCCAG	356
QY	827	PheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGlyLysAspGln	846
Db	357	TAAAGAAAGTACACACAAATAACCCAGACGAACTCGATCAAAACAGGGGAAAAACCA	416
QY	847	HisPheProValPheMetAsnGluLysGluAspIleLeuTyrCysThrGluMetGluArg	866
Db	417	CTTTTCCCTGCTTGATATATGCAAGAAATGTTNGTGGTACACTGAGCTCGAAGG	476
QY	867	ValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeuAlaArgGln	886
Db	477	NTCTTTGGCTTTCTCTGTGACACTACACAGACGTGTCCAAATGGGCGCTGTGCGCCGAC	536
QY	887	ArgLeuLeuGlyArgSerTyrPheSerValProValIleArgHisLeuPheAlaProLeuLys	906
Db	537	AAGCTGCTGGGAAAGTCTGAGCGTGCTGTATCCAGACACTTTCGCCCTCTGAAG	596
QY	907	GluTyrPheAlaCys	911
Db	597	GACTACTTGCATGT	611

**THIS PAGE BLANK (USPTO)**

[illegible]

**THIS PAGE BLANK (USPTO)**

```

1177 ACTCGTGGGTGATGTTGTTGCGAGATGCGAACTTCTCAGTGTGTGTGTGGAGAACTC 1236
QY 341 MetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnPromet 360
DB 1237 ATGCCGCTGAGCTCTCTCTCTGAGTGCATTCCACAGGCCACTCAACAAAGAGCCCATG 1296
QY 361 TyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeuPhe 380
DB 1297 TACCGCAAGCCATCAACGAGTCTCCAGGTGGCCAGCCGCGCGGAAGCTGTTT 1356
QY 381 ProAlaCysHisAspSerAspGluSerAspSerGlyLysAlaValGluValGlnAsnLys 400
DB 1357 CCAGCTTGGCAGTACAGTGTGTAAGTGAAGTGAAGTGTGGAAGTGAAGTGAAGTGAAG 1416
QY 401 GlnMetIleGluTyrAlaLeuGluGlyPheGlnProSerGlyProLysGlyLeuGluPro 420
DB 1447 CAGATGATTAATGGGCTCTGAGTGTCTCCAGCCCTCGGCTCAAGGCGCTGGAGCCA 1476
QY 1477 ProGluGluGluLysAsnProTyrLysGluValTyrThrAspMetTyrValGluProGlu 440
DB 1477 CCAGAAAGAGAGAGAAATCTTACAGAGAAATTTACCCGACATGGGTGGAGCTGAA 1536
QY 441 AlaAlaAlaTyrAlaProProProAlaLysLysProArgLysSerThrThrGluLys 460
DB 1537 GCAGCTGCTTACGCCCAACCCCAAGCAGCAAGAAACCCAGAAAGAGCAACACAGAGAA 1596
QY 461 ProLysValLysGluLysIleAspGluArgThrArgGluArgLysValTyrGluValArg 480
DB 1597 CCTAAGGTCAGAGAGATCATTGATGAGCGCACAGGAGCGGCTGTGTATGAGTGGTGGC 1656
QY 481 GlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThrLeu 500
DB 1657 CAGAAAGTCAGAAACATCGAGACATTTGATCTCATGTGGAGCTTCAATGTCAACCTTG 1716
QY 501 GlnHisProLeuPheIleGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGluCys 520
DB 1717 GAGACCCACTCTTCTTATGGAGGCACTGCGCAACTGTAAAGACTGCTTCTTGGAGTGT 1776
QY 521 AlaTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyGlyArg 540
DB 1777 GCTTAACCACTATGACCAAGATGGATGATACCACTTATGACCACTTGTGCTGGGGGGCT 1836
QY 541 GluValLeuMetCysGlyAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsnAsn 560
DB 1837 GAAGTGTCTATGTGTGGAAACAACAACCTGCTGCAAGTGTCTTGTGTCAAGTGTGTGAT 1896
QY 1897 LeuLeuValGlyProGlyAlaAlaAlaAlaAlaIleLysGluAspProTyrAsnCysTyr 580
DB 1897 CTCTTGGTGGGGCCAGAGAGCTGCTTCAAGCACTTAAAGAAAGACCTTGGAACTGCTAC 1956
QY 581 MetCysGlyHisLysGlyThrTyrGlyLeuLeuArgArgGluAspTyrProSerArg 600
DB 1957 ATGTCCGGGCAATAGGCACTATGGCTGCTGCGAAAGCGGAAAGACTGGCTTCTCGA 2016
QY 601 LeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProLysValTyrPro 620
DB 2017 CTCCAGATGTCTTGTGCAATACCATGACCAAGAAATTTGACCCCCCAAGGTTTACCCA 2076
QY 621 ProValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIleAla 640
DB 2077 CCTGTGCCAGCTGAGAAAGAGAGAGCCATCCGCTGCTGTCTCTTTCATGGGATTGCT 2136
QY 641 ThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSerGlu 660
DB 2137 ACAGGCTCTCTGTCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2196
QY 661 ValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyrVal 680
DB 2197 GTGTGTGAGAGACTCCATCAGCTGGGCACTGTCGGCAAGAGAGAGAGAGAGAGAGAGAG 2256
QY 681 GlyAspValArgSerValThrGlnLysHisIleGlnGluTyrGlyProPheAspLeuVal 700
DB 2257 GGGAGCTGCGGAGCTCACACAGAGCATATCCAGAGTGGGGCCCATTTGACCTGTGTG 2316

```

```

QY 701 IleGlyLysSerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeuTyr 720
DB 2317 ATTGAGGCACTCTCTGCAATGACCTTCCATTGTCTAACCTTGGCCGCAAGGAGACTTAT 2376
QY 721 GlnGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgProLys 740
DB 2377 GAGGGTCTGGCGGCTCTTCTTGTGAGTTTACCGCTCTCTGCATGATGCGGCGCCAG 2436
QY 741 GlnLysAspAspArgProPhePheTyrPheGluGluAsnValValAlaMetGlyValSer 760
DB 2437 GAGGAGATGATGCGCCCTCTTCTGTGGCTCTTGAAGATGTGTGGCATAGGCGCTTGTAG 2496
QY 761 AspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLysGlu 780
DB 2497 GACAGAGGACATCTCGGATTTCTTGAGTCTAACCCCTGTATGATTTGAGCCCAAGAA 2556
QY 781 ValSerAlaAlaHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMetAsnArgPro 800
DB 2557 GTGTCTGTGACACAGGCGCTTACTTCTGGGTTAACCCTTCTGCGATGACAGCGCT 2616
QY 801 LeuAlaSerThrValAsnAspLysLeuGluLeuGlnGlyCysLeuGlnHisGlyArgIle 820
DB 2617 TTGGCATCCACTGTGAATGATAGCTGAGCTGCAAGAGTGTGAGAGAGAGAGAGAGAGAG 2676
QY 821 AlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGlyLys 840
DB 2677 GCCAAGTTCACAAAGTGAAGACCATTAACCAAGCTCAAACTTATTAAGCAGGCGCAA 2736
QY 841 AspGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTyrCysThrGlnMet 860
DB 2737 GACCAAGATTTCCCGCTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2796
QY 861 GluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeuAla 880
DB 2797 GAAAGGCTGTGGCTTCCCGTCCACTACACAGAGCTCTCCAAATGAGCGCTGGCG 2856
QY 881 ArgGlnArgLeuLeuGlyArgSerTyrSerValProValIleArgHisLeuPheAlaPro 900
DB 2857 AGGCAAGAGACTGCTGGCGGATCGTGAAGCGTCCGCTCATTCGCCCACTTCTGCTCCG 2916
QY 901 LeuLysGluTyrPheAlaCysVal 908
DB 2917 CTGAAGAAATATTTGTGTGTG 2940

```

RESULT 2

**THIS PAGE BLANK (USPTO)**

## RESULT

AAT21884

ID AAT21884 standard; cDNA to mRNA; 301 BP.

XX

AC AAT21884;

XX

DT 14-AUG-1996 (first entry)

XX  
DE Human gene signature HUMGS03426.XX  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.XX  
OS Homo sapiens.XX  
PN WO9514772-A1.XX  
PD 01-JUN-1995.XX  
PF 11-NOV-1994; 94WO-JP01916.XX  
PR 12-NOV-1993; 93JP-0355504.XX  
PA (MATS/) MATSUBARA K.

PA (OKUB/) OKUBO K.

XX  
PI Matsubara K, Okubo K;XX  
DR WPI; 1995-206931/27.XX  
PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissuesXX  
PS Claim 1; Page 991-992; 2245pp; Japanese.XX  
CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in AAT19001-T26837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.XX  
SQ Sequence 301 BP; 96 A; 65 C; 63 G; 66 T; 11 other;Query Match 6.6%; Score 285.2; DB 16; Length 301;  
Best Local Similarity 95.3%; Pred. No. 1.5e-48;  
Matches 287; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 3519 GATCAGATAGGAGCACAAGCAGGGGACGGAAAGAGAGAGACACTCAGGCGGCAGCATTCC 3578  
 Db 1 GATCAGATAGGAGCACAAGCAGGGGACGGAAAGAGAGAGNCACTCANGCGGCAGCATTCC 60

QY 3579 CTCCAGCCACTGAGCTGTCGTGCCAGCACCATTCTCTGGTCACGCAAAACAGAACCCAGT 3638  
 Db 61 CTNCCAGCCACTGAGCTGNCGTGCCAGCACCATTCTCTGGTCACGCAAAACAGAACCCATT 120

QY 3639 TAGCAGCAGGGAGACGAGAACCACACAAAGACATTTTCTACAGTATTTTCAAGTGCCTA 3698  
 Db 121 TAGCAGCAGGGAGACGAGAACCACACAAAGACATTTTCTACAGTATTTTCAAGTGCCTA 180

QY 3699 CCACACAGGAAACCTTGAAGAAAATCAGTTTCTAGAAGCCGCTGTACCTCTTGTTTACA 3758  
 Db 181 CCACACAGGAAACCTTGAAGAAAATCAGTTTCTAGGNNNGCTGTACCTCTTGTTTACA 240

QY 3759 GTTTATATATATATGATAGATATGAGATATATATATAAAAGGTAAGTAACTACTGTAC 3818  
 Db 241 GTTTATATATATATGATAGATATGNGTATATATATAAAAGGTAAGTAACTACTGTAA 300

QY 3819 A 3819  
Db 301 A 301



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 22, 2002, 03:58:32 / Search time 5173.7 Seconds

(without alignments)  
5107.633 Million cell updates/sec

Title: US-09-720-086-5

Perfect score: 4909

Sequence: 1 MPSSGPGTSSSLEREDDR.....MSVPIRHLPAKKEFACV 908

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool/US09720086/runat.18112002.092832.22195/app.query.fasta.1.4252  
-DB=genembl -OPMT=fastac -SUFFIX=rge -MINMATCH=0.1 -LOOFCU=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09720086.qcgn.1.11308@runat.18112002.092832.22195 -NCPU=6 -ICPU=3  
-NO XLPXY -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_scs:\*  
28: em\_un:\*

29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_ey:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4909	100.0	4192	10	AF068625 Mus muscu
2	4903	99.9	4094	10	BC007466 Mus muscu
3	4729.5	96.3	3005	9	AF067972 AF067972 Homo sapi
4	4729.5	96.3	4258	9	AF331856 AF331856 Homo sapi
5	2571.5	52.4	123936	2	AC120824 AC120824 Rattus no
6	2412	49.1	2191	9	AK025230 AK025230 Homo sapi
7	2221.5	45.3	168651	9	AC009474 AC009474 Homo sapi
8	2211	45.0	4135	10	AF068627 AF068627 Mus muscu
9	2211	45.0	4278	10	AF151970 AF151970 Mus muscu
10	2206	44.9	4338	10	AF068626 AF068626 Mus muscu
11	2206	44.9	4338	10	AF151969 AF151969 Mus muscu
12	2203.5	44.9	4163	10	AF151974 AF151974 Mus muscu
13	2200.5	44.8	176697	2	AC112586 AC112586 Rattus no
14	2200	44.8	4233	10	AF151973 AF151973 Mus muscu
15	2198.5	44.8	4233	10	AF151973 AF151973 Mus muscu
16	2195.5	44.7	4267	9	AF156488 AF156488 Homo sapi
17	2187	44.6	4145	9	AF156488 AF156488 Homo sapi
18	1938.5	39.5	3946	10	AF068628 AF068628 Mus muscu
19	1938.5	39.5	4089	10	AF151971 AF151971 Mus muscu
20	1933.5	39.4	4006	10	AV078427 AV078427 Mus muscu
21	1933.5	39.4	4149	10	AF151972 AF151972 Mus muscu
22	1931	39.3	3974	10	AF151976 AF151976 Mus muscu
23	1926	39.2	4034	10	AF151975 AF151975 Mus muscu
24	1908.5	38.9	3897	9	AF156487 AF156487 Homo sapi
25	1835.5	37.4	204534	2	AC118195 AC118195 Mus muscu
26	1835.5	37.4	208874	5	AF135438 AF135438 Danio rer
27	1763.5	35.9	2848	5	AF135438 AF135438 Danio rer
28	1748.5	35.6	197652	9	AC012074 AC012074 Homo sapi
29	1623.5	33.1	3017	9	AK001191 AK001191 Homo sapi
30	1504	30.6	1758	9	BC018214 BC018214 Homo sapi
31	1445.5	29.4	210269	2	AC116459 AC116459 Mus muscu
32	1184	24.1	185413	2	AC119009 AC119009 Rattus no
33	1032.5	21.0	119630	2	AC112040 AC112040 Rattus no
34	996.5	20.3	167568	2	AC111734 AC111734 Rattus no
35	994	20.2	2127	9	AF129267 AF129267 Homo sapi
36	872.5	17.8	2057	9	AF129268 AF129268 Homo sapi
37	785	16.0	2008	9	AF129269 AF129269 Homo sapi
38	748	15.2	2077	6	AR129189 AR129189 Sequence
39	728.5	14.8	119630	2	AC112040 AC112040 Rattus no
40	692.5	14.1	1397	9	BC002560 BC002560 Homo sapi
41	688	14.0	1705	9	AF194032 AF194032 Homo sapi
42	644.5	13.1	1589	10	AF220524 AF220524 Mus muscu
43	644.5	13.1	1645	10	AF220524 AF220524 Mus muscu
44	561.5	11.4	244329	2	AC107644 AC107644 Mus muscu
45	490.5	10.0	204534	2	AC118195 AC118195 Mus muscu

#### ALIGNMENTS

RESULT 1

AF068625 4192 bp mRNA linear ROD 06-DEC-1999  
 LOCUS Mus musculus DNA cytosine-5 methyltransferase 3A (Dnmt3a) mRNA,  
 complete cds.  
 ACCESSION AF068625 GI:6449467  
 VERSION AF068625.2  
 KEYWORDS Mus musculus.  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 4192)  
 Okano.M., Xie.S. and Li.E.  
 Cloning and characterization of a family of novel mammalian DNA  
 (cytosine-5) methyltransferases  
 Nat. Genet. 19 (3), 219-220 (1998)  
 MEDLINE 98324766  
 PUBMED 9662389  
 2 (bases 1 to 4192)  
 Xie.S., Okano.M. and Li.E.  
 Direct Submision  
 Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,  
 Charlestown, MA 02129, USA  
 3 (bases 1 to 4192)  
 Okano.M., Chijiwa.T., Sasaki.H. and Li.E.  
 Direct Submision  
 Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,  
 Charlestown, MA 02129, USA  
 Sequence update by submitter  
 On Nov 18, 1999 this sequence version replaced gi:3327977.  
 FEATURES  
 Location/Qualifiers  
 1..4192  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosomes="12"  
 /map="4.0 cM"  
 1..4192  
 /gene="Dnmt3a"  
 1..4192  
 /gene="Dnmt3a"  
 /note="similar to EST sequences deposited in GenBank  
 Accession Numbers AA052791, AA111043, AA154890, AA240794,  
 AA756653, W58898, W59299, W91664, and W91665"  
 217..2943  
 /gene="Dnmt3a"  
 /function="de novo DNA methylation"  
 /note="contains Cys-rich region"  
 /codon\_start=1  
 /product="DNA cytosine-5 methyltransferase 3A"  
 /protein\_id="AAC40177.2"  
 /db\_xref="GI:6449468"  
 /translation="MPSSGPGDTSSSSLERDDRRKEGEQEEENRKEEROBPSATARK  
 VGRPKRKHPPVESDTPKPAVTTKSPMAQDSGPSDLLPNGDLKRSRPOEES  
 PAQQRKGAPAGEGTEPEASRAVENGCVTKEGRGASAGEKEKQTNIESMKWE  
 GSRGLRGLGWESSLRQRPRLTFOAGDPYITISKRKDLNLRKREAEKKAIVIA  
 VNNAVEENQASGESQKVEASPPAVQPTDPSPVATTPPEPVGGDAGDKNAIKADD  
 EPEYEDRGFGIGELVMGKLRFSWMPGRIVSWMTGRSRAAGLTFVWFGDGSFV  
 VCVEKLPLSFCFAHQATYKQPMYKAIYEVIVASRAGLTFVWFGDGSFV  
 KAVEVNOMIEMALGCFQSPGKLEPBEENKPYKEVYTDMMVEPEAAAYAPPPA  
 MKPKSTTKPKVKEIIDETRELVYEVOKCRNIEDICISCSLNVTLHPLEFIG  
 MCQCNKCLECAYQDDGQYQVCTICCGGREVLNMCNNCCRCFCVCCVDLLVPGP  
 AQAATKEDPWCYMGHGKTYGLLRREDWPSRLQMFANNHQBEPDPKVPVPPA  
 ERKPTIRLSLFGIATGLLVLDLGTQVDYITASEVCEOSITVGMVRHQKIMYVD  
 KSRVTKRIQEWGPFDLVIGSNCNLSIYNPARKGLYEGTGRLEFFRYLLHDARPK  
 EGDRLFFWLFEVNVAVGVSKRDISRFLSNPVMIDAKESAAHRRARYFNGNLPQWN  
 RPLASTVNDKLEQCLEHGRIFAKFSKRTITTSNSIKQSKDQHPFVFMKEEDILW  
 CTEMERVFPGPHYTDVSNMSRLARQLLGRSVSPVIRHLFAPLKEYFACV"  
 BASE COUNT 1094 a 1072 c 1186 g 840 t  
 ORIGIN  
 Alignment Scores: 3 25e-211 Length: 4192  
 Pred. No.: 4909.00 Matches: 908  
 Score:

Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 10 Gaps: 0  
 US-09-720-086-5 (1-908) x AF068625 (1-4192)  
 Qy 1 MetProSerSerGlyProGlyAspThrSerSerSerSerLeuGluArgGluAspAspArg 20  
 Db 217 ATGCCCTCCAGCGGCCCGGGACACACAGCAGCTCTCTCTGAGCGGGAGGATGATCGA 276  
 Qy 21 LysGluGlyGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 40  
 Db 277 AAGGAAGGAGAGGAACAGGAGGAGAACCTGTCGACGAGGAGGAGGAGGAGGAGGAGG 336  
 Qy 41 ThrAlaArgLysValGlyArgProGlyArgLysArgLysArgLysArgLysArgLys 60  
 Db 337 ACGGCCCGGAAGGTGGGAGGCTTGGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGT 396  
 Qy 61 AspThrProLysAspProAlaValThrLysSerGlnProMetAlaGlnAspSerGly 80  
 Db 397 GACACCCCAAGGAGCCAGCAGTGCACCAAGTCTCAGCCCATGGCCAGGACTCTGGC 456  
 Qy 81 ProSerAspLeuProAsnGlyAspLeuGluLysArgSerGluProGlnProGluGlu 100  
 Db 457 CCCTCAGATCTCTACCCCAATGGAGACTTGGAGAGCGGAGTGAACCCCAACCTGAGGAG 516  
 Qy 101 GlySerProAlaAlaGlyGlnLysGlyGlyAlaProAlaGluGlyGluGlyThrGluThr 120  
 Db 517 GGGAGCCCAAGTTCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 576  
 Qy 121 ProProGluAlaSerArgAlaValGluAsnGlyCysValThrLysGluGlyArgGly 140  
 Db 577 CCACACAGAGCCTCCAGAGCTGTGGAGAAATGGCTGTGTGTGTGTGTGTGTGTGTGTGA 636  
 Qy 141 AlaSerAlaGlyGluGlyLysGluGlnLysGlnThrAsnLysGluSerMetLysMetGlu 160  
 Db 637 GCCTCTGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 696  
 Qy 161 GlySerArgGlyArgLeuArgGlyGlyLeuGlyTrpGluSerSerLeuArgGlnArgPro 180  
 Db 697 GGCCTCCCGGGCGGAGCTGGCAGGTGGCTGGGCTGGGAGTCCAGCCTCCGTCAGGAGCCC 756  
 Qy 181 MetProArgLeuThrPheGlnAlaGlyAspProTyrTyrLysSerLysArgLysArgAsp 200  
 Db 757 ATGCCAAGACTCACCTTCCAGGCGAGGGGACCCCTACTACATCAGCAACCGGAAACGGAT 816  
 Qy 201 GluTrpLeuAlaArgTrpLysArgGluAlaGluLysLysAlaLysValLysAlaValMet 220  
 Db 817 GAGTGTGTCACGTCGTGGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 876  
 Qy 221 AsnAlaValGluGluAsnGlnAlaSerGlyGluSerGlnLysValGluGluAlaSerPro 240  
 Db 877 AATGCTGTGCAGAGAACCAAGGCTCTGGAGAGTCTCAGAAAGGTGGAGGAGGAGGAGGAG 936  
 Qy 241 ProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrThrProGluPro 260  
 Db 937 CCTGCTGTGCAGAGCCCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 996  
 Qy 261 ValGlyLysAspAlaGlyAspLysAsnAlaThrLysAlaAlaAspAspGluProGluTyr 280  
 Db 997 GTAGGAGGAGGATCTGGGGACAAAGATGCTACCAAGAGCAGCGAGGAGGAGGAGGAGGAGT 1056  
 Qy 281 GluAspGlyArgGlyPheGlyLysGlyGluLeuValTrpGlyLysLeuArgGlyPheSer 300  
 Db 1057 GAGGATGGCCCGGGCTTTGGCATTTGGAGAGCTGCTGTGGGGGAAACTTTCGGGGCTTCTCC 1116  
 Qy 301 TrpTrpProGlyArgLysValSerTrpMetThrThrGlyArgSerArgAlaAlaGluGly 320  
 Db 1117 TGTGGCCAGCCCGAATTGTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1176  
 Qy 321 ThrArgTrpValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLysLeu 340

Db 1177 ACTCGCTGGATCATGTGTTCCGAGATGCAAGATTCTCAGTGTGTGTGGAGAGCTC 1236  
 QY 341 MetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnProMet 360  
 Db 1237 ATGCCCGTGAAGCTCTTCTGCACTGTCATTCCACAGGCCACCTAACAGAGCCCATG 1296  
 QY 361 TyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeuPhe 380  
 Db 1297 TACCGCAAGGCATCTACCAAGTCTCCAGGTGGCCAGCAGCCGTGCCGGGAAGCTGTTT 1356  
 QY 381 ProAlaCysHisAspSerAspGluSerAspSerGlyValAlaValGluValGlnAsnLys 400  
 Db 1357 CCAGCTTGCCCTGACAGTATGAAGTGCAGTGGCAGAGCTGTGGAAGTGCAAGCAAG 1416  
 QY 401 GlnMetIleGluTTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGluPro 420  
 Db 1417 CAGATGATTTGAATGGGCCCTCGGTGGCTTCCAGCCCTCGGCTCTTAAGGGCTGGAGCCA 1476  
 QY 421 ProGluGluGluLysAsnProTyrTyrGluValTyrThrAspMetTrpValGluProGlu 440  
 Db 1477 CCAGAGAGAGAGAGAAATCTTTCAGAGAGTTTACACCGACATGTGGTGGAGCTGAA 1536  
 QY 441 AlaAlaAlaTyrAlaProProProAlaLysLysProArgLysSerThrThrGluLys 460  
 Db 1537 GCACTGCTTACGCCCCACCCCAAGCAAGCAAGAAAGAGAGAGAGAGAGAGAGAA 1596  
 QY 461 ProLysValLysGluIleIleAspGluArgThrArgGluArgLeuValTyrGluValArg 480  
 Db 1597 CCTAAGGTCAAGAGATCATTTGATGAGGAGCAAGAGGAGCGGTGTATGAGAGGCGC 1656  
 QY 481 GlnLysCysArgAsnIleGluAspIleCysArgLysSerCysLysLeuAsnValThrLeu 500  
 Db 1657 CAGAGAGCAAGAAATCGAGGACATTTGTATCTCATGTGGAGCTCAATCTCACCTCG 1716  
 QY 501 GlnHisProLeuPheIleGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGluCys 520  
 Db 1717 GAGCACCACCTCTTATTGGAGGATGTGCCAGAACTGTAAAGAACTGTTCTTGGAGTGT 1776  
 QY 521 AlaTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyArg 540  
 Db 1777 GCTTACCAAGTATGAGAGCAATGTGGTACCAGTCTTATTTGACCACTGTGTGGGGGGC 1836  
 QY 541 GluValLeuMetCysGlyLysAsnAsnAsnCysCysArgCysPheCysValGluCysValAsp 560  
 Db 1837 GAAGTGTCTCATGTGTGGAGCAACAACTGTGAGGTGCTTTGTGTGCGAGGTGTGAT 1896  
 QY 561 LeuLeuValGlyProGlyValAlaAlaGlnAlaIleLysGluAspProTrpAsnCysTyr 580  
 Db 1897 CTCTTGTGGGGCCAGAGCTGCTCAGGCAAGCATTAAAGAGAAAGCCCTGGAAGCTGTAC 1956  
 QY 581 MetCysGlyHisLysGlyTyrTyrGlyLeuLeuArgArgArgGluAspTrpProSerArg 600  
 Db 1957 ATGTGCCGGGCACTAAGGGCACCTATGGCTGTGCGAAGACGGGAAGACTGGGCTTCTCGA 2016  
 QY 601 LeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProLysValTyrPro 620  
 Db 2017 CTCACAGATGTTCTTTGCCAATTAACATGACCAAGAAATTTGACCCCAAGAGTTTACC 2076  
 QY 621 ProValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIleAla 640  
 Db 2077 CCTGTGCCAGCTGAGAGAGAGAGAGCCCATCCGCTGTCTCTCTTTGATGGGATTTGCT 2136  
 QY 641 ThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAlaAspArgTyrIleAlaSerGlu 660  
 Db 2137 ACAGGGCTCTGTGTCTGAAGAGACTGGGCACTCCAGTGAACCCCTCAATTCCTCCGAG 2196  
 QY 661 ValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyrVal 680  
 Db 2197 GTGTGTAGGACTCCATCACGCTGGGCACTGTGGCGGCAAGGAAAGATCATGTACGTC 2256  
 QY 681 GlyAspValArgSerValThrGlnLysHisIleGlnGluTyrGlyProPheAspLeuVal 700  
 Db 2257 GGGGACGTCCGACGCTCACACAGAGCATATCCAGAGTGGGGCCCATTCGACCTGTGTG 2316

QY 701 IleGlyGlySerProCysAsnAspLeuSerIleValAlaAsnProAlaArgLysGlyLeuTyr 720  
 Db 2317 ATTTGAGGGCGATCCCTGCAATACCTCTTCATTGTCAACCCCGCCCAAGGAGCTTAT 2376  
 QY 721 GluGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgProLys 740  
 Db 2377 GAGGGTACTGGCCCGCTCTTCTTTGAGTTCTTACCGCTCTCTCATATGCGGGGCCAAG 2436  
 QY 741 GluGlyAspAspArgProPhePheTrpLeuPheGluAsnValValAlaMetGlyValSer 760  
 Db 2437 GAGGGAGATATCGCCCTTCTTCTGCTCTTTAGAAATGTGTGGCCCAATGGGCGTTAGT 2496  
 QY 761 AspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLysGlu 780  
 Db 2497 GACAAAGAGGACATCTCCGATTTCTTGAGCTAACCCCGTATGATTGACGCCAAGAA 2556  
 QY 781 ValSerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMetAsnArgPro 800  
 Db 2557 GTGTCTGCTCACACAGGGCCGTTACTTGTGGGGTAACTTCTGCGCATGAACAGGCCT 2616  
 QY 801 LeuAlaSerThrValAsnAspLysLeuGluLeuGlnGlyCysLeuGluHisGlyValArg 820  
 Db 2617 TTGGCATCCACTGTGAATGATAGCTGAGCTGCAGAGAGTGTCTGGAGCAGCGCAATA 2676  
 QY 821 AlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGlyLys 840  
 Db 2677 GCCAAGTTCAGCAAGAGAGGACCAATTAACCAAGGTCAAACTCTTAAGCAGGGCAAA 2736  
 QY 841 AspGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTyrCysThrGluMet 860  
 Db 2737 GACCAGCATTTCCCGCTTCTTCAATGAACGAGAGAGGACATCTGTGTGTGACTGAAGA 2796  
 QY 861 GluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeuAla 880  
 Db 2797 GAAAGGTGTTTGGCTTCCCGCTCACTACACAGAGCGTCTCCAAATGAGCCGCTTGGG 2856  
 QY 881 ArgGlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAlaPro 900  
 Db 2857 AGGCAGAGACTGCTGGGCCGATCTGTGAGGCTGCGGTATCCGCCACTTCTGCTCCG 2916  
 QY 901 LeuLysGluTyrPheAlaCysVal 908  
 Db 2917 CTGAAGGAATTTTGTGTGTG 2940  
 RESULT 2  
 BC007466 4094 bp mRNA linear ROD 07-AUG-2002  
 LOCUS  
 DEFINITION Mus musculus, DNA methyltransferase 3A, clone MGC:5662  
 IMAGE:3492853, mRNA, complete cds.  
 ACCESSION BC007466  
 VERSION BC007466.1 GI:13938620  
 KEYWORDS MGC.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 4094)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-MAY-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 CONTACT: MGC help desk  
 Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
 Tissue Procurement: Gilbert Smith, Ph.D.  
 CDNA Library Preparation: Life Technologies, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULNL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human  
 Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www.shgc.stanford.edu>

Contact: (Dickson, Mark) mcd@paxil.stanford.edu  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAK Plate: 8 Row: n Column: 11  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6681208.

## FEATURES

source

1..4094  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /map="CZEBCH 11"  
 /clone="MGC:5662 IMAGE:3492853"  
 /tissue\_type="Mammary tumor metastasized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."  
 /clone\_lib="NCI CGAP\_Lu29"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6"  
 245..2971

CDS

/codon\_start=1  
 /product="DNA methyltransferase 3A"  
 /protein\_id="AAH07466.1"  
 /db\_xref="GI:13938621"  
 /db\_xref="LocusID:13435"  
 /translations="MPSSSGPQDTSSSLERDDRRKEGEQENRGKEROBPSATARK  
 VGRPKRKHPPVESDTPKDAVTKSOPMAQDSGPSDLLPNDLDEKRSRPOPEGS  
 PAAGOKGAPAGEGETETPEASRAVENGCCVTKEGASAGEKEOKPTNLESMKE  
 GGRGLRGLGHESSLRQRPMPLTFQAGDPYIYIKRDEWLARWREAEKAKVIA  
 VNVAENQASGESQVKEASPPAQQPTDFASPTVAITPEFGDGDGDKNATKAADD  
 BEYEDRGFGICELVWGLRFSWPGRIYVWMTGRSRAEGRWVMWFGDGKFSV  
 VCEKMLPSSFCFSAHQATYKQPKYKAIYEVLQVASSRAGLFFACHDSDESFG  
 KAVEQNMKIEWALGGFQSPKGLPEEKEKNPKYEVYTDMMVEPEAAAYAPPPA  
 KPKRSTKPKVKEIIDETRELVYVQKCRNIEDICISCSGLNVTLLHPLFIGG  
 MCQCNKCFLEKAYQYDDGYQYCTICCGREVLGCMNNNCRFCVCEVDLLVPGG  
 AQAALKEPWPNCMGHGTGLYLRREDWPSRLQMFANNHQDEPPPKVYPPVPA  
 EKRPRLVSLFDGIATGLVLKGLQVDRYIASEVEDSITVGMVPHQCKIMYVG  
 VSVTKHLOEAGPPFDLVIGSSPCNDLSIVNPAKGLYEGTGRLLFFRYRLHDA  
 EGDRFPFLFENVVANGSDKADISRFLESNPNVDAKEVSAHRAIRYFNGNLPFM  
 RELATNDKLEQCEHRIKAKSKVRTITTSNSIKQSGDHPFVFMKEKIDILW  
 CTEMERVFVGFPHYTDVSNMSRLARQLRLLGRSMVVPVIRHLFAPLKEYFACV"

BASE COUNT 1067 a 1057 c 1164 g 806 t

## Alignment Scores:

Pred. No.: 5,9e-211 Length: 4094  
 Score: 4903.00 Matches: 907  
 Percent Similarity: 99.89% Conservatives: 0  
 Best Local Similarity: 99.89% Mismatches: 1  
 Query Match: 99.88% Indels: 0  
 DB: 10 Gaps: 0

US-09-720-086-5 (1-908) x BC007466 (1-4094)

QY 1 MetProSerSerGlyProGlyAspThrSerSerSerSerLeuGluArgGluAspArg 20  
 Db 245 ATGCCCTCCAGCGGCCCCGGGACCCAGACAGCTCTCTTGAGCCGGAGGATGATCA 304

QY 21 LysGluGlyGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 40  
 Db 305 AAGGAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 364

QY 41 ThrAlaArgLysValGlyArgProGlyArgLysArgLysHisProProValGluSerSer 60  
 Db 365 ACGGCCCGGAAGTGGGAGGCTGGCGGGAAGCGAAGCACCCACCGGTGGAAGCAGT 424

QY 61 AspThrProLysAspProAlaValThrThrLysSerGlnProMetAlaGlnAspSerGly 80  
 Db 425 GACACCCCAAGACCCAGCAGTACCAAGTCTCAGCCATGGCCACGACTCTCGC 484

QY 81 ProSerAspLeuLeuProAsnGlyAspLeuGluLysArgSerGluProGluGlu 100

Db 485 CCCTCAGATCTGCTACCAATGGAGACTTGGAGAGCGGAGTGAACCCCACTGAGGAG 544

QY 101 GlySerProAlaAlaGlyGlnLysGlyGlyAlaProAlaGluGlyGluGlyThrGluThr 120  
 Db 545 GGGAGCCCACTGTCAGGGCAGAGGGTGGGCCGCCAGCTGAAGGAGAGGAGAACTGAGACC 604

QY 121 ProProGluAlaSerArgAlaValGluAsnGlyCysCysValThrLysGluGlyArgGly 140  
 Db 605 CCACCAGAAGCCTCCAGAGCTGTGGAGATGGCTGCTGTGTGACCAAGGAGGCGCTGGA 664

QY 141 AlaSerAlaGlyGluGlyLysGluGlnLysGlnThrAsnIleGluSerMetLysMetGlu 160  
 Db 665 GCCTCTCAGAGAGAGGCGCAAGAACAGAGCGCACCAATCGAATCCATGAAAAATGGAG 724

QY 161 GlySerArgGlyArgLeuArgGlyGlyLeuGlyTrpGlySerSerLeuArgGlnArgPro 180  
 Db 725 GGCTCCCGGGGCCACTGCGAGGTGGCTGGCTGGAGTCCAGCTCCGCTCAGCGACCC 784

QY 181 MetProArgLeuThrPheGlnAlaGlyAspProTyrTyrIleSerLysArgLysArg 200  
 Db 785 ATGCCAAGACTCACCCTCCAGGAGGGAGCCCTACTACTACAGCAACGGAACCGGAT 844

QY 201 GluTrpLeuAlaArgTrpLysArgGluAlaGluLysLysAlaLysValIleAlaValMet 220  
 Db 845 GAGTGGCTGCGCACCTTGGAAAAAGGAGGCTGAGAGAAAGCAAGGTAAATTGCAGTAATG 904

QY 221 AsnAlaValGluGluAsnGlnAlaSerGlyGluSerGlnLysValGluGluAlaSerPro 240  
 Db 905 AATGCTGTGGAAGAGAACCCAGGCTCTGGAGAGTCTCAGAGGTGGAGGAGGAGCCCTT 964

QY 241 ProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrThrProGluPro 260  
 Db 965 CCTGCTGTGAGCAGCCAGCCAGGACCTCTCTCCGACTGTGGCCACCAACCCCTGAGCCA 1024

QY 261 ValGlyGlyAspAlaGlyAspLysAsnAlaThrLysAlaAlaAspAspGluProGluTyr 280  
 Db 1025 GTAGGAGGGATGCTGGGACCAAGATGCTACCAAGCAGCCGAGCATGAGCTGAGTAT 1084

QY 281 GluAspGlyArgGlyPheGlyLleGlyGluLeuValTrpGlyLysLeuArgGlyPheSer 300  
 Db 1085 GAGATGCCCGGGCTTTGGCATTTGGAGAGCTGGTGGGGGAAACTTCGGGGCTTCTCC 1144

QY 301 TrpTrpProGlyArgLleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGluGly 320  
 Db 1145 TGGTGGCCAGCGCGCAATTGTGCTTTGGTGGATGACAGGCGGAGCCGAGCAGCTGAAGGC 1204

QY 321 ThrArgTrpValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLysLeu 340  
 Db 1205 ACTGCTGGGTCTAGTGGTTCGGAGATGGCAAGTTCTCAGTGGTGTGTGTGGAGAAAGCTC 1264

QY 341 MetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnProMet 360  
 Db 1265 ATGCCGCTGAGCTCTTCTGCAAGTGCATTCCACAGGCGCACCTACAACAAGCAGCCCATG 1324

QY 361 TyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeuPhe 380  
 Db 1325 TACCGCAAGCCATCTACGAAAGTCTCCAGGTGGCCAGCCGCTGCCGGAGCTGTTTT 1384

QY 381 ProAlaCysHisAspSerAspGluSerAspSerGlyLysAlaValGluValGlnAsnLys 400  
 Db 1385 CCAGCTTGCCATGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1444

QY 401 GlnMetIleGluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGluPro 420  
 Db 1445 CAGATGATTCAATGGGCGCTCGGTGGCTTCCAGCCCTCGGTCTTAAGGGCTTGAACCA 1504

QY 421 ProGluGluLysAsnProTyrLysGluValTyrThrAspMetTrpValGluProGlu 440  
 Db 1505 CCAGAGAGAGAGAGAAATCTTTACAAGAAAGTTTACACCGACATGTGGTGGAGCTGTGA 1564

QY 441 AlaAlaAlaTyrAlaProProProAlaLysLysProArgLysSerThrThrGluLys 460

Db	1565	GCAGGTGCTTAACGCCCCACCCCAACCAAGCAAGAAACCCAGAAAGACCAACAAGAGAAA	1624
QY	461	ProlysvallysgluileleaspjiuargThrargjiuargleuValTyrGluValArg	480
Db	1625	CCTAAGTCAAGAGATCATTCATGATGACCGCAACAGGAGCGGCTGGTGTATGAGAGTGGCC	1684
QY	481	GlnlysvsArgshniiegluaspiliecysileserCysgliserLeuasnValThrleu	500
Db	1685	CAGAAGTCAGAAACATCGAGACATTTGTATTCATGTCGGAGGACCTCAATGTCAACCTG	1744
QY	501	GluHisProleupheileglYglYmetCysglinsncysylsasncysPheleuglucys	520
Db	1745	GAGCAACCACTCTTCATTGGAGGCACTGTCAGAACTGTAAAGAACTGCTTGGAGGT	1804
QY	521	AlatyrGlnTyrAspaspaspIYtyrGlnserTyrCysThrIleCysCysglYglYarg	540
Db	1805	GCTTACCAAGTATACAGACGATGGTACACTCTTATTCACCACTCTGTCGGGGGGCGT	1864
QY	541	GluValleuMetCysglYasnasnancysCysArgCysPheCysValGluCysValasp	560
Db	1865	GAAAGTCTATATGTGTGGGAACAACAACCTGCTGCAGGTGCTTTGTGTGAGTGTGTGAT	1924
QY	561	LeuleuValGlyProGlyValAlaAlaGlnAlaAlaIlelysgluaspProTyrAsnCysTyr	580
Db	1925	CTCTTGTGTGGGCCAGAGACTGCTCAGCGACCACTTAAAGMAAGACCCCTGGAACTGTAC	1984
QY	581	MetCysGlyHislysglyThrtYtyrGlyLeuleuArgArgGluaspTyrProserArg	600
Db	1985	ATGTGCGGCATAAAGGCACTATGAGCTGTGCGAAGCGGGAAGACTGGCTTCTCGA	2044
QY	601	LeuGlnMetPhePheAlaAsnAsnHisaspGlnGluPheAspPropolysValTyrPro	620
Db	2045	CTCCAGATGTTCTTGGCAATACCATGACCAGGAATTTGACCCCCCAAGGTTTAAACCA	2104
QY	621	ProValProAlaGluLyArgLyAspPoleargValleuSerleuPheaspGlyIleAla	640
Db	2105	CCTGTGCCAGCTGAGAGAGAGAGGCCATCCGCGTGTCTCTCTTTGATGGGATTGCT	2164
QY	641	ThrGlyLeuleuValleuLyAspLeuGlyIleGlnValAspArgTyrIleAlaSerGlu	660
Db	2165	ACAGGACTCTGTGTGCTGAAGACCTGGGCACTCCAAAGTGAACCGCTACATTGCCCTCCAG	2224
QY	661	ValCysGluaspserIleThrValGlyMetValArgHisglnglyLysIleMetTyrVal	680
Db	2225	GTGTGTGAGACTTCATACGGGTGGCATGTGTCGGCACACGAGAAAGATCATGTACGTC	2284
QY	681	GlyAspValArgSerValThrGlnLysHisIleGlnGlnTyrProPheaspLeuVal	700
Db	2285	GAGGACGTCCGCAAGCTACACAGAAACATATCCAGAGTGGGGCCCATTCGACCTGGTG	2344
QY	701	IleGlyGlySerProCysAsnaspLeuSerIleValAsnProIleArglyGlyLeuTyr	720
Db	2345	ATTGAGAGCACTCCCTGCAATACCTCTCCATTGTCAACCTCCGCCCAAGGACCTTAT	2404
QY	721	GluGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgProlys	740
Db	2405	GAGGTACTGGCGCCCTCTCTTCTTGAAGTTCTACCGACTCTCTGATGATGCCGGCCCAAG	2464
QY	741	GluGlyAspAspArgProPhePheTyrPhePheGluAsnValValAlaMetGlyValSer	760
Db	2465	GAGGAGATGATCGCCCTCTCTTCTGAGCTCTTGAAGAAAGTGGGACCATGGGCGTTAGT	2524
QY	761	AspLyAspArgaspIleSerArgPheleuGluSerAsnProValMetIleaspAlaLyGlu	780
Db	2525	GACCAAGAGGACACCTCGCGAATTTCTTGAGCTTAAACCCGTGATGATGACGCCAAAGAA	2584
QY	781	ValSerAlaAlaHisArgAlaArgTyrPheTyrProGlyAsnLeuProGlyMetAsnArgPro	800
Db	2585	GTGTGCTGCACACAGGCGCCGTTATCTTGGGGTTAACTTCTTGATGAAACAGGCT	2644
QY	801	LeuAlaSerThrValAsnAspLyLeuGlnGlnGlyLeuGlnHisglYArgIle	820
Db	2645	TTGGCAATCACTGTGAATGATTAAGCTGAGAGTGTGCAAGAGAGTGTGAGACACGCGAATA	2704

AGDDEPYEDGGRGFIGELVWGLKIRFESWPGRIIVSWWNTGSRRAAEGTRVWVWFGDQ  
KFSVCVEMKMLPSFSAEGHQATYKNQPMYKRAIVELYOVASSRAGKLVPVCHDSDE  
SDTAKAVEQNQMPSIALGQFGSPGQEPPEKEENKIVKEYVTOMVYVEPEAAAYH  
PPAKPRKSTAEKPKVEI1DDERTYRVLVYVQROKRIED1C8GSLNTEHLEPL  
VGMWCNKNCFKVEI1DDERTYRVLVYVQROKRIED1C8GSLNTEHLEPL  
VGPCAAQAA1KEDPMNVCYMGCHGKYGALLREDDPSRLQMPFNHNDQEDFDPKVPV  
VPAKXKPIRVLSIFDGIATGLLVKLDQIOVDNYIASEVCEDSITVGNVRHQGKIM  
YVGVRSYQKHIOEMGFDDVLVIGSPGNDLISVPAKGLGETGLRFPFVRLHLD  
AGNRRDPRPFWFLFENAGHVRKIDRPLESNPMV1DAKESAAHRAVFWGNL  
PMNRP2P2P2VNDKELECHGR1AKFSKRYTIT1TNSN1KQGDQHPFVPMFNE  
DILWCTBERVFGFPVHYDTSN507 t 846 c 978 c 507 t

BASE COUNT	674 a	846 c	978 a	507 c
BASE COUNT	674 a	846 c	978 a	507 c

Alignment Scores:

Alignment Scores:					
Read No.:	2.48e-203	Length:	3005		
Score:	4729.50	Matches:	875		
Percent Similarity:	97.14%	Conservative:	8		
Best Local Similarity:	96.26%	Mismatches:	25		
Query Match:	96.34%	Indels:	1		
DB:	9	Gaps:	1		

US-09-720-086-5 (1-908) x AF067972 (1-3005)

Qy	1	MetProSerSerGlyProGlyAspThrSerSerSerSerLeuGluArgGluAspAspArg	20
Db	239	ATGCCCTCCAGCGGCCCGGGGACACAGCAGCTCTGCTCGGAGCGGAGAGGACCGA	298
Qy	21	LysGluGlyGluGluGlnGluAsnArgGlyLysGluGluArgGlnGluProSerAla	40
Db	299	AAGGACGGAGAGGAGCAGGAGGAGCGCGTGGCAAGGAGGAGCGCCAAAGAGCCAGCACC	358
Qy	41	ThrAlaArgLysValGlyArgProGlyArgLysArgLysHisProProValGluSerSer	60
Db	359	ACGGCACGAAGGTGGCGCGCTGGGAGGAAGCGCAAGCACCCCCCGTGGTGAAGCGGT	418
Qy	61	AspThrProLysAspProAlaValThrThrLysSerGlnProMetAlaGlnAspSerGly	80
Db	419	GACACGCCAAAGSACCCTGGCGGTGATCTCCAAGTCCCCATCCATGGCCAGGACTCAGGC	478
Qy	81	ProSerAspLeuLeuProAsnGlyAspLeuGluLysArgSerGluProGlnProGluGlu	100
Db	479	GCCTCAGAGCTATTACCCAATGGGACTTGGAGAAGCGGAGCTGAGCCCCCAGCCAGGAG	538
Qy	101	GlySerProAlaAlaGlyGlnLysGlyGlyAlaProAlaGluGlyGluGly---ThrGlu	119
Db	539	GGGAGCCCTGCTGGGGGGCAGAAAGCGCGGGCCCCCAGCAGAGGGAGAGGGTGCAGCTGAG	598
Qy	120	ThrProProGluAlaSerArgAlaValAlaGlnGlyCysCysValThrLysGluGlyArg	139
Db	599	ACCCCTGCCTGAAGCCTCAAGAGCAGTGGAAAAATGGCTGCTGCACCCCAAGAGGCGCCA	658
Qy	140	GlyAlaSerAlaGlyGluGlyLysGluGlnLysGlnThrAsnIleGluSerMetLysMet	159
Db	659	GGAGCCCTGCAGACGGGCAAGAAGAACAGAGAGACCAACATCGAATCCATGAANAATG	718
Qy	160	GluGlySerArgGlyArgLeuArgGlyGlyLeuGlyTrpGluSerSerLeuArgGlnArg	179
Db	719	GAGGCTCCCGGGCGCGCTCGCGGGTGGGTGGGCTGGGASTCCAGCTCCGTCAGCGG	778
Qy	180	ProMetProArgLeuThrPheGlnAlaGlyAspProTyrTrpIleSerLysArgLysArg	199
Db	779	CCCATGCCGAGGCTCACCTTCCAGCGGGGGGACCCCTACTACATCAGCAAGCGCAAGCGG	838
Qy	200	AspGluTrpLeuAlaArgTrpLysArgGluAlaGluLysLysAlaLysValIleAlaVal	219
Db	839	GACGAGTGGCTGGCAGCTTGAAGAAGGGAGGCTGAGAAGAAGCCACAGGTCAATGCAGA	898
Qy	220	MetAsnAlaValGluGluAsnGlnAlaSerGlyGluSerGlnLysValGluGluAlaSer	239
Db	899	ATGAATGCTGTGGAAGAAAACCAAGGGGCCCGGGGAGTCTCAGAAGGTGCAGAGGCCACG	958
Qy	240	ProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrThrProGlu	259







Db 527 GGGAGCCCTGCTGGGGGCGAAGAGGCGGGGCCCCAGCAGAGGAGAGGGTGCAGCTGAG 586  
Qy 120 ThrProGluAlaSerArgAlaValGluAsnGlyCysValThrLysGluGlyArg 139  
Db 587 ACCCTGCTTGAAGCCTCAAGAGCAGTGGAAAAATGGCTGTCACCCCAAGAGAGCGCGA 646  
Qy 140 GlyAlaSerAlaGlyGluGlyLysGluGlnLysGlnThrAsnIleGluSerMetLysMet 159  
Db 647 GGAGCCCTGCAGAGCGGCAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 706  
Qy 160 GluGlySerArgGlyArgGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 179  
Db 707 GAGGGCTCCCGGGGCGGCTGCGGGTGGCTGGGGCTGGGAGTCCAGCTCCGTCAGCGG 766  
Qy 180 ProMetProArgLeuThrPheGlnAlaGlyAspProTyrIleSerLysArgLysArg 199  
Db 767 CCCATGCCAGGCTCACCTTCCAGGCGGGGAGCCCTACTACATCAGCAAGCGCAAGCGG 826  
Qy 200 AspGluTTPLeuAlaArgTTPlysArgGluAlaGluLysLysAlaLysValIleAlaVal 219  
Db 827 GACGAGTGGCTGGCAGCGCTGAAAGAGGAGGCTGAGAAAGAAAGCAAGGTCAATTGCAGGA 886  
Qy 220 MetAsnAlaValGluGluAsnGlnAlaSerGlyGluSerGlnLysValGluGluAlaSer 239  
Db 887 ATGAATGCTGTGAAAGAAACCAAGGCGGGGAGTCTCAGAAAGTGGAGAGGCCAGC 946  
Qy 240 ProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrProGlu 259  
Db 947 CCTCTGCTGTGCAGCAGCCCACTGACCCCGCATCCCGCATCTGGCTACCCAGCGCTGAG 1006  
Qy 260 ProValGlyGlyAspAlaGlyAspLysAsnAlaThrLysAlaAlaAspAspGluProGlu 279  
Db 1007 CCGGTGGGTCCGATGCTGGGGACAAATGCCAACCAAGCAGGCGATCAGCAGGCCAGAG 1066  
Qy 280 TyrGluAspGlyArgGlyPheGlyIleGlyGluLeuValTTPGlyLysLeuArgGlyPhe 299  
Db 1067 TACGAGAGCGCGGGCTTTGGCATTTGGGAGAGTGGTGTGGGGGAAATCTCGGGGGCTTC 1126  
Qy 300 SerTTPProGlyArgIleValSerTTPTrpMetThrGlyArgSerArgAlaAlaGlu 319  
Db 1127 TCTGTGTGCAGCGCCGCAATGTGTCTGTGTGGATGACGGGCGGAGCGGAGCAGCTGAA 1186  
Qy 320 GlyThrArgTTPValMetTTPPheGlyAspGlyLysPheSerValValCysValGluLys 339  
Db 1187 GGCACCCGCTGGGTTCATGTGTTCCGAGACGGCAATTTCTCAGTGGTGTGTGGAGAG 1246  
Qy 340 LeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnPro 359  
Db 1247 CTGATCCGCTGAGCTGTTTGGTGTGCGTTCACACAGGCCACCTACAAACAGCAGCCC 1306  
Qy 360 MetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeu 379  
Db 1307 ATGTACCGCAAGCCATCTACGAGGTCTGCAAGGTGGCAGCAGCGCGCGGGAAGCTG 1366  
Qy 380 PheProAlaCysHisAspSerAspGluSerAspSerGlyLysAlaValGluValGlnAsn 399  
Db 1367 TTCCCGGTGTCACAGCAGCGATGAGAGTGACATGCCAAGGCGGTGGAGGTGCAGAAAC 1426  
Qy 400 LysGlnMetIleGluTTPAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGlu 419  
Db 1427 AAGCCCATGATTGAATGGGCTCTGGGGGCTTCCAGCCCTTCTGGCCCTAAGGGCCTGGAG 1486  
Qy 420 ProProGluGluLysAsnProTyrLysGluValTyrThrAspMetTTPValGluPro 439  
Db 1487 CCACCAAGAAAGAGAAATCCCTACAAAGAAAGTGTACACGACATGTGGGTGGAACTT 1546  
Qy 440 GluAlaAlaIleTyrAlaProProAlaLysLysProArgLysSerThrThrGlu 459  
Db 1547 GAGCAGCTGCTACGCAACCACTCCACCAAGCAAAAGCCCGGAAGAGCAGCGGAG 1606  
Qy 460 LysProLysValLysGluIleAlaAspGluArgThrArgGluLeuValTyrGluVal 479

Db 1607 AAGCCCAAGGTCAAGGAGATTATTGATGAGCGCACAGAGAGCGGCTGGTGTACGAGGTG 1666  
Qy 480 ArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThr 499  
Db 1667 CGGCAGAAAGTGGCGGAACATTCAGGACATTCGATCTCTCTGTGGAGCCTCAATGTTTACC 1726  
Qy 500 LeuGluHisProLeuPheIleGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGlu 519  
Db 1727 CTGGAAACACCCCTCTTCTGTGGAGAAATGTGCMAAACTGCAGAACTGCTTTCTGGAG 1786  
Qy 520 CysAlaTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyGly 539  
Db 1787 TGTGCGTACACAGTACGACGACGCGCTACAGTCTCTACTGCACCATCTGTGTGGGGG 1846  
Qy 540 ArgGluValIleLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysVal 559  
Db 1847 CGTGAAGTCTCATGTGCGGAACAACTGCTGCGAGGTCTTTTGTGCGAGTGTGTG 1906  
Qy 560 AspLeuLeuValGlyProGlyAlaAlaGlnAlaIleLysGluAspProTTPAsnCys 579  
Db 1907 GACCTCTTGGTGGGCGGGGCTGCCCGAGGACCATTAAGGAAGACCCCTGGAACTGC 1966  
Qy 580 TyrMetCysGlyHisLysGlyThrTyrGlyLeuLeuArgArgArgGluAspTTPProSer 599  
Db 1967 TACATGTGCGGCGACAAGGGTACTTACCGGCTGCTGCGCGCGAGAGACTGGCGCTCC 2026  
Qy 600 ArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProLysValTyr 619  
Db 2027 CGSCTCCAGATGTTCTTCTGCTAATACACGACCGAATTTGACCTTCCAAAGGTTTAC 2086  
Qy 620 ProProValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIle 639  
Db 2087 CCACCTGTCCAGCTGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATC 2146  
Qy 640 AlaThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSer 659  
Db 2147 GCTACAGGCTCTCTGTGCTGAGGACTTGGGCATTGAGGTGGACCGCTACATTCCTCG 2206  
Qy 660 GluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyr 679  
Db 2207 GAGGTGTGTAGAGCTCCATCACGTTGGCATGTGTGGGCACAGGGGAAAGATCATGTAC 2266  
Qy 680 ValGlyAspValArgSerValThrGlnLysHisIleGlnGluTTPGlyProPheAspLeu 699  
Db 2267 GTCGGGGACGTCGCGAGCTCACAGAGCATATCCAGGAGTGGGGGCCCATTCGATCTG 2326  
Qy 700 ValIleGlyGlySerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeu 719  
Db 2327 GTGATTGGGGGAGTCCCTGCATGACCTCTCCATCTGTCACCCCTGCTCGAAGGGCCCTC 2386  
Qy 720 TyrGluGlyThrGlyArgLeuPheGluPheTyrArgLeuLeuHisAspAlaArgPro 739  
Db 2387 TACGAGGACGTCGCGGCTCTTCTTTGAGTTCTACCGGCTCTGCAATGATCGCGGCC 2446  
Qy 740 LysGluGlyAspAspArgProPhePheTTPLeuPheGluAsnValAlaMetGlyVal 759  
Db 2447 AAGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2506  
Qy 760 SerAspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLys 779  
Db 2507 AGTGACAAAGAGGAGATCTCGCATTTCTCGAGTCCAAACCCCTGTGATGATGATGATGAT 2566  
Qy 780 GluValSerAlaAlaHisArgAlaArgTyrPheTTPGlyAsnLeuProGlyMetAsnArg 799  
Db 2567 GAGGTGTCAGCTGACACAGGCGCGCTTCTTCTGGGTAACTTCTCCCGGTATGAACAGG 2626  
Qy 800 ProLeuAlaSerThrValAsnAspLysLeuGluLeuGlnCysLeuGluHisGlyArg 819  
Db 2627 CCGTGGCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2686  
Qy 820 IleAlaLysPheSerLysValArgThrIleThrArgSerAsnSerIleLysGlnGly 839  
Db 2687 ATAGCCCAAGTTTCAAGAAAGTGAAGGACCATTTACTAGAGGTCAAACTCCATAAAGCAGG 2746

QY 840 LysAspGlnHisPheProValPheMetAsnGluValGluAspIleLeuTyrCysThrGlu 859  
 DB 2747 AAGACACGACGATTTCCTCTTCATGATGAAGAAGACATCTTATGAGCACTGAA 2806  
 QY 860 MetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeu 879  
 DB 2807 ATGGAAGGGGATTGTTGTTCCAGTCACATACTAGAGCTTCACATGAGCGCTTG 2866  
 QY 880 AlaArgGlnArgLeuLeuGluValArgSerTyrPheSerValProValIleArgHisIlePheAla 899  
 DB 2867 GCGAGGAGAGAGACTGCTGGCCGCTCATGAGCGCTGCAGTCATCCGACCTCTTCGCT 2926  
 QY 900 ProLeuValGluTyrPheAlaCysVal 908  
 DB 2927 CCGCTGAAGAGATTTCGCTGTGTG 2953  
 REF 5  
 AC120824  
 LOCUS 123936 bp DNA linear HTG 23-JUL-2002  
 DEFINITION Rattus norvegicus clone CH230-282D16, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\*, 38 unordered pieces.  
 AC120824  
 AC120824.2 GI:21902541  
 HTG: HTG, PHASE1.  
 KEYWORDS Rattus norvegicus.  
 SOURCE Rattus norvegicus.  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 123936)  
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amaralunga,H.C., Are,J.R., Ayale,M., Banks,T.,  
 Barberia,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,  
 Bouck,J., Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P.,  
 Buhay,C., Butch,P., Burkett,C., Butrell,K.L., Byrd,N.C.,  
 Carion,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escott,M.,  
 Falle,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
 Gabisti,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Gortell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hognes,M., Hollway,C., Hollins,B.,  
 Homsl,F., Howard,S., Huber,J., Huiyk,S., Hume,J., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,  
 Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,B.,  
 Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,  
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
 Nguyen,N., Nickerson,E., Nwokwenkwo,S., Ogub,M., Okunnu,G.,  
 Orangunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Picken,R., Primus,B., Pu,L.L., Qulies,M., Ren,Y.,  
 Rives,M., Rojas,A., Rojubenkan,I., Rolfe,M., Ruiz,S., Savery,G.,  
 Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,  
 Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,  
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Umaney,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
 Williams,G., Williamson,A., Wleczkyk,R., Wooden,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstein,G. and Gibbs,R.  
 TITLE Direct Submission  
 JOURNAL Unpublished

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT  
 2 (bases 1 to 123936)  
 Worley,K.C.  
 Direct Submission  
 Submitted (13-MAY-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 123936)  
 Worley,K.C.  
 Direct Submission  
 Submitted (23-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 18, 2002 this sequence version replaced gi:20531783.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GMLV  
 Center clone name: CH230-282D16  
 ----- Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 95907 bases at least Q40  
 Consensus quality: 99223 bases at least Q30  
 Consensus quality: 101036 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length.  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 38 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \*  
 1 1009: contig of 1009 bp in length  
 1010 1109: gap of unknown length  
 1110 2113: contig of 1004 bp in length  
 2114 2213: gap of unknown length  
 2214 3233: contig of 1020 bp in length  
 3234 3333: gap of unknown length  
 3334 4442: contig of 1109 bp in length  
 4443 4542: gap of unknown length  
 4543 6202: contig of 1660 bp in length  
 6203 6302: gap of unknown length  
 6303 7464: contig of 1162 bp in length  
 7465 7564: gap of unknown length  
 7565 8772: contig of 1208 bp in length  
 8773 8872: gap of unknown length  
 8873 10371: contig of 1499 bp in length  
 10372 10471: gap of unknown length  
 10472 11714: contig of 1243 bp in length  
 11715 11814: gap of unknown length  
 11815 12924: contig of 1110 bp in length  
 12925 13024: gap of unknown length  
 13025 14347: contig of 1323 bp in length  
 14348 14447: gap of unknown length  
 14448 15652: contig of 1205 bp in length  
 15653 15796: gap of unknown length  
 15797 17797: gap of 2044 bp in length  
 17798 17896: gap of unknown length  
 17897 18937: contig of 1041 bp in length  
 18938 19037: gap of unknown length  
 19038 21543: contig of 2506 bp in length  
 21544 21644: gap of unknown length  
 21645 24540: contig of 2897 bp in length  
 24541 24640: gap of unknown length  
 24641 28112: contig of 3472 bp in length  
 28113 28212: gap of unknown length

\* 28213 31101: contig of 2889 bp in length  
\* 31102: gap of unknown length  
\* 34415: contig of 3214 bp in length  
\* 34416: gap of unknown length  
\* 37353: contig of 2838 bp in length  
\* 37354: gap of unknown length  
\* 37454: contig of 3485 bp in length  
\* 40339: gap of unknown length  
\* 44331: contig of 3293 bp in length  
\* 44332: gap of unknown length  
\* 44432: contig of 3604 bp in length  
\* 48035: gap of unknown length  
\* 48036: gap of unknown length  
\* 48135: contig of 3046 bp in length  
\* 51181: gap of unknown length  
\* 51182: contig of 4080 bp in length  
\* 55361: gap of unknown length  
\* 55362: contig of 4119 bp in length  
\* 59580: gap of unknown length  
\* 59581: contig of 3845 bp in length  
\* 63525: gap of unknown length  
\* 63526: contig of 3809 bp in length  
\* 67434: gap of unknown length  
\* 67435: contig of 4660 bp in length  
\* 72194: gap of unknown length  
\* 72195: contig of 5383 bp in length  
\* 72295: gap of unknown length  
\* 77677: contig of 3572 bp in length  
\* 77778: gap of unknown length  
\* 81349: contig of 4947 bp in length  
\* 81350: gap of unknown length  
\* 8450: contig of 4728 bp in length  
\* 86397: gap of unknown length  
\* 86497: contig of 6445 bp in length  
\* 91324: gap of unknown length  
\* 91325: contig of 7801 bp in length  
\* 97769: gap of unknown length  
\* 97770: contig of 5598 bp in length  
\* 97870: gap of unknown length  
\* 105670: contig of 5212 bp in length  
\* 105671: gap of unknown length  
\* 113369: contig of 7156 bp in length  
\* 113369: gap of unknown length  
\* 11469: contig of 7156 bp in length  
\* 116681: gap of unknown length  
\* 116781: contig of 7156 bp in length.

## FEATURES

Location/Qualifiers  
1..123936  
/organism="Rattus norvegicus"  
/db\_xref="taxon:10116"  
/clone="CH230-282D16"

BASE COUNT 35904 a 21413 c 21468 g 36486 t 8665 others

## Alignment Scores:

Pred. No.: 3.62e-105 Length: 123936  
Score: 2571.50 Matches: 487  
Percent Similarity: 92.22% Conservative: 23  
Best Local Similarity: 88.07% Mismatches: 27  
Query Match: 52.38% Indels: 18  
DB: 2 Gaps: 4

US-09-720-086-5 (1-908) x AC120824 (1-123936)

Qy 203 LeuAlaArgTrpLysArgGluAlaGluLysAlaLysValIleAlaValMetAsnAla 222  
Db 84783 CTGGCCCTATGG-----GTTGAGAGAAAGCCCAAGGTGATTGCGAGTAATCAATGCT 84833  
Qy 223 ValGluAlaAsnGlnAlaSerGlyGluSerGlnLysValGluAlaSerProProAla 242  
Db 84834 GTGGAGGGAACACGACCTCTGGGAGTCTCAGAAAGTGGAGCGCCCTCTCTGCA 84893  
Qy 243 ValGlnProThrAspProAlaSerProThrValAlaThrProGluProValGly 262  
Db 84894 GTGCAGCAGCCAG-GACCCCTGCATCTCCACCTGTGGCCATTACCCCTGTGCCAGTAGG 84952  
Qy 263 GlyAspAlaGlyAspLysAsnAlaThrLysAlaAlaAspAspGluProGluLysAsp 282

Db 84953 GCTGATGCTGGGACAGAATGCTATCAAGGCAGCTGATGAGGACCCCGAGTATGAGGAT 85012  
Qy 283 GlyArgGlyPheGlyIleGlyGluLeuValTrpGlyLysLeuArgGlyPheSerTrpTrp 302  
Db 85013 GGCAGGCGCTTGGCATTTGGAGAGCTGGTGTGGGGAAGAACTCGGGGGCTTCTCCTGGTGG 85072  
Qy 303 ProGlyArgIleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGluGlyThrArg 322  
Db 85073 CCAAGCCGAATTGTCTTGGTGGATGACAGCCAGAGCAGAGCAGCCGAAAGCACTCGC 85132  
Qy 323 TrpValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLysLeuMetPro 342  
Db 85133 TGGGTCACTGTGGTTTGGAGATGGCAAGTTTTCAGTGATGTGTGCGAAGAGCTCATGCC 85192  
Qy 343 LeuSerSerPheCysSerAlaPheHisGlnAlaThrTrpAsnLysGlnProMetTrpArg 362  
Db 85193 CTGAGCTCTTCTGCAT-----TTCACCCAGGCCACCTCAACAAGCAGCCCATGTACCCG 85248  
Qy 363 LysAlaIleTrpGluValLeuGlnValAlaSerSerArgAlaGlyLysLeuPheProAla 382  
Db 85249 AAGCCATCTACGAAGTCTCTCAGGTGCCAGCAGCTGTTCAGGGAAGCTGTTCGCCAGCG 85308  
Qy 383 CysHisAspSerAsp--GluSerAspSerGlyLysAlaValGluValGlnAsnLysGlnM 402  
Db 85309 TGTATGACAGCGGACATGAAGTGCACACTGGCAAGCTGTGGAGGTGCAGAAACAGAGA 85368  
Qy 402 etIleGluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGluProProG 422  
Db 85369 CGATTGAGTGGTCTTGGCGGCTTCCAGCCCTCTGGTCCCAAGGCGCTGGAGCCACCAG 85428  
Qy 422 LuGluGluLysAsnProTrpLysGluValTrpThrAspMetTrpValGluProGluAla 442  
Db 85429 AAGAGGAGAACATCTTTACAGGAGTGTACCCAGCATGTGGGTGGAGCCCTGAGGCAG 85488  
Qy 442 LaAlaTrpAlaProProProAlaLysLysProArgLysSerThrThrGluLys-Pro 461  
Db 85489 CTGCTTATTCCTCCACCCACCCAGCCAGCAAGAAACCCAGAAAGAGC-----ACAACCT 85539  
Qy 462 LysValLysGluIleIleAspGluArgThrArgGluArgLeuValTrpGluValArgGln 481  
Db 85540 AAGGTCAAGGAGATCATTTGATGAACACACAGAGAACCCACTCATGTATGAGGTGCCAG 85599  
Qy 482 LysCysArgAsnIleGluAlaApIleCysIleSerCysGlySerLeuAsnValThrLeuGlu 501  
Db 85600 AAGTGCACAAACATCGAGGACATTTGTATCTCATGTGGAGCCTCAATGTCACTCCCTGGAG 85659  
Qy 502 HisProLeuPheIleGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGluCysAla 521  
Db 85660 CACCCACTCTTCGTTGGTGAATGTGCCAGAACTGTAAAGAACTGCTTCTTGGAGTGTGCT 85719  
Qy 522 TrpGlnTrpAspAspAspGlyTrpGlnSerTrpCysThrIleCysCysGlyGlyArgGlu 541  
Db 85720 TACCAATACCATGACGATGTGTACAGTCTCTTGTACCATCTGTCTGGGGGCCCGAA 85779  
Qy 542 ValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysValAspLeu 561  
Db 85780 GTGCTCATGTGTGGGAACAACACCGCTGCAGGTGCTTTTGTGTGGAGTGTGTGGATCTC 85839  
Qy 562 LeuValGlyProGlyAlaAlaGlnAlaAlaIleLysGluAspProTrpAsnCysTrpMet 581  
Db 85840 TTGGTGGGGCCAGGGCTGCCCAAGCAGCCATTAAAGGAAGACCCCTCGGAACCTGCTACATG 85899  
Qy 582 CysGlyHisLysGlyThrTrpGlyLeuLeuArgArgGluAspTrpProSerArgLeu 601  
Db 85900 TGT-----GAGCTGCTGGGAGAGAGGAGGAGTGGCCCTTCTCGACTC 85941  
Qy 602 GlnMetPheAlaAsnAsnHisAspGlnGluPheAspProProLysValTrpProPro 621  
Db 85942 CAGATGTTCTTTCACATAACCCAGCAGCAAGAAATTTGACCCCGGAAAGGTTTACCACCT 86001  
Qy 622 ValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIleAlaThr 641  
Db 86002 GTGCCAGCTGAGAGAGAGAGCCCATCCAGGTGTATCTTCTATTGATGGGATGCTGCTACA 86061

```

OY 642 G1yleu1eVal1leu1yAsp1e1eG1y1leG1n1Val1Asp1aG1T1T1le1a1SeG1u1a1 661
Db 86062 GGGCTCTGGTGGCTGAAGAAGCACTGGGCGATCCAAAGTGAAGCCCTACATTGCTTCGAGGG 86121
OY 662 CysG1u1Asp1Ser11eTh1r1Val1G1yMe1t1Ala1Gh1SeG1n1y1y1S1eMe1T1y1Val1G1y 681
Db 86122 TGCGAGACATCCATCAACCGTGGGATAGTGCAGCACCAAGGAGGAGGAGGATCATGTTGTCGGG 86181
OY 682 Asp1Ala1Arg1Ser1a1Th1r1G1n1y1Sh1e1G1n1G1u1T1r1p1y1r1P1r1o1p1h1e1Asp1Leu1Val1le 701
Db 86182 GACGCTGCAGCATCACACAGAAAGCATATCCAGAGAGGGGGCCCATACATCATCTGTGTGATT 86241
OY 702 G1yG1y1Ser1Pro1Cy1As1Asp1Leu1Ser11eVal1Asp1Pro1a1a1r1G1y1Se1y1Leu1T1r1G1u 721
Db 86242 GGGGCGAGTCTTTCGAATGATCACTCTCCATCATCAACCTCGCCGCGCAAGGACATTACAAAG 86301
OY 722 G1yTh1G1y1a1r1G1e1r1P1r1o1p1h1e1G1u1P1h1e1T1r1a1r1G1e1u1H1a1s1p1a1a1a1r1P1r1o1y1Se1y1 741
Db 86302 GGCACCTGGCGACCTCTTCTTGTGAAGTTCAACCACTCTCGATGATGCGCGCCGAAGAG 86361
OY 742 G1yAsp1Asp1a1r1G1r1P1r1o1p1h1e1T1r1P1r1e1u1h1eG1u 752
Db 86362 GGAGAGATCGCCCTCTTCTGTGCTCTTTGAG 86394
RESULT 6
AKO25230
LOCUS
DEFINITION
Homo sapiens cDNA: FLJ21577 fis, clone COL06724, highly similar to
AF067972 Homo sapiens DNA cytosine methyltransferase 3 alpha
(DNMT3A) mRNA.
ACCESSION
AKO25230
VERSION
AKO25230.1 GI:10437699
KEYWORDS
oligo capping; fis (full insert sequence).
SOURCE
Homo sapiens colon cDNA to mRNA, clone_1ib:COL clone:COL06724.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Carnitina; Homiinda; Homo.
REFERENCE
1 (sites)
Kawabata,A., HikiJi,T., Kobatake,N., Inagaki,H., Ikema,Y.,
Okamoto,S., Okitani,R., Oca,T., Suzuki,Y., Obayashi,M., Nishi,T.,
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2191)
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human
Genome Center, Shirokane-dai, 4-6-1, Minato-Ku, Tokyo 108-8639,
Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
Location/Qualifiers
1..2191
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="COL06724"
/tissue_type="colon"
/clone_1ib="COL"
/note="cloning vector pME18SFLJ3"
1..2191
/note="highly similar to AF067972 Homo sapiens DNA
cytosine methyltransferase 3 alpha (DNMT3A) mRNA"
BASE COUNT 571 a 505 c 575 g 540 t
ORIGIN

```

Alignment Scores:	
Prod. No.:	8_3e-100
Score:	2412.00
Percent Similarity:	100.00%
Best Local Similarity:	99.55%
Query Match:	49.13%
DB:	9
	Gaps: 0
	Indels: 0
	Mismatches: 0
	Conservative: 2
	Matches: 440
	Length: 2191

US-09-720-086-5 (1-908) X AK025230 (1-2191)

OY		467	IleAspGluArgThrArgGluValTyrGluValAlaArgGlnLysCysArgAsnIle	486
Dd		2	ATTATGAGCGCAACAAGACGGCTGGTGTAAGAAGTGGGCAGAAATGCCGAACATT	61
OY		487	GluAspIleCysIleSerCysGlySerLeuAsnValThreGluHisProLeuPheIle	506
Dd		62	GAGACATCTGCATCTCTGTGGAGCTCAATGTAACTCTGGAAACCCTCTTCGTT	122
OY		507	GlyGlyMetCysGlyAsnCysLysAsnCysPheLeuGluCysAlaTyrGluNTyrAspAsp	526
Dd		122	GGAGGAATGTGCCAANAACCTCACAAACTGCTTTCTGGAGTGTGGCTACAGAACGAC	189
OY		527	AspGlyTyrGlnSerTyrCysThrIleCysCysGlyGlyValArgGluValLeuMetCysGly	546
Dd		182	GACGGCTACAGTCTCACTGACCATCTGCTGTGGGGCCGCTAGGTGCTCATGTGGGA	241
OY		547	AsnAsnAsnCysCysArgCysPheCysValGluCysValAspLeuLeuValGlyProGly	566
Dd		242	AACAACAACCTGCTGCAGGTGCTTTTGGCTGAGTGTGGAACTCTTGTTGGGGCCGGGG	300
OY		567	AlaAlaGlnAlaAlaIleLysGluAspProTyrPasnCysTyrMetCysGlyHisLysGly	586
Dd		302	GCTGCCACGGCAGCATTTAAGGAAGACCCCTGGAACTGTCAATGTGGCGGCAACAGGCT	360
OY		587	ThrTyrGlyLeuLeuArgArgArgGluAspTyrProSerArgLeuGlnMetPheAla	600
Dd		362	ACCTAAGGGCTGTGGCGGGCGAGAGAGACTGCCCTCCGGCTCCAGATGTCTTCGCT	422
OY		607	AsnAsnHisAspGlnGluPheAspProProLysValTyrProProValProAlaGluLys	626
Dd		422	AATAACCAACGACCAAGAAATTGACCTCCAAGGTTTACCACCTGTCCCAGCTGAGAG	483
OY		627	ArgLysProIleArgValLeuSerLeuPheAspGlyIleAlaThrGlyLeuLeuValLeu	646
Dd		482	AGGAAGCCCATCCGGGTGCTGTCTCTTTGATGATATGCCTACAGGGCTCCTGTGGCTGG	541
OY		647	LysAspLeuGlyIleGlnValAspArgTyrIleAlaSerGluValCysGluAspSerIle	666
Dd		542	AAGGACTTGGGCAATTGAGGTGGACCGGCTACATTGCCCTGGAGGGTGTGAGAGACTTCATC	600
OY		667	ThrValGlyMetValArgHisGlnGlyLysIleMetTyrValGlyAspValArgSerVal	686
Dd		602	ACGGTGGGCAATGGTGGCGCACCAAGGGAAGATCATGTACGTGGGGAGCTCCGCAAGCTC	666
OY		687	ThrGlnLysHisIleGlnGluTyrGlyProPheAspLeuValIleGlyLysSerProCys	706
Dd		662	ACACGAAGCATATCCAGAGATGGGGGCCCATTCGATCTGTGATTTGGGGGAGTCCCTGC	722
OY		707	AsnAspLeuSerIleValAsnProAlaArgLysGlyLeuTyrGluGluTyrGlyArgLeu	726
Dd		722	AATGACCTCTCCATCGTCAACCTGCTGGCAAGGGCTCTTACGAGGGCACTGGCCGGCTC	783
OY		727	PhePheGluPheTyrArgLeuLeuHisAspAlaArgProLysGluGlyAspAspArgPro	746
Dd		782	TTCCTTGAGTTCTACCGCCCTCTGCTGATATGGCGGGCCCAAGAGAGGAGATATCCGCC	844
OY		747	PhePheTrpLeuPheGluAsnValAlaIleMetGlyValSerAspLysArgAspIleSer	766
Dd		842	TTCCTCTGGCTCTTTGAGATGTGGTGGCAATGGGGTTTAGTGAACAAGAGCAATCTCG	900
OY		767	ArgPheLeuGluSerAsnProValMetIleAspAlaLysGluValSerAlaAlaHisArg	786



```
repeat_region /rpc_family="CR1"
9304..9469
/rpc_family="Alu"
repeat_region /rpc_family="Alu"
10317..10439
/rpc_family="MIR"
repeat_region /rpc_family="Alu"
12215..12455
/rpc_family="Alu"
repeat_region /rpc_family="L1"
12467..13537
/rpc_family="L1"
repeat_region /rpc_family="L1"
13553..14453
/rpc_family="L1"
repeat_region /rpc_family="L1"
14853..15123
/rpc_family="MER2_type"
16209..16345
/note="similar to BST AA338793 (NID:G1991103)"
misc_feature /note="similar to BST AA338793 (NID:G1991103)"
16209..16345
/note="similar to BST AA1123298 (NID:G539064) ov48c08.x1"
misc_feature /note="similar to BST AA1123298 (NID:G539064) ov48c08.x1"
16210..16345
/note="similar to BST AL037015 (NID:G5928008)"
repeat_region /note="similar to BST AL037015 (NID:G5928008)"
16513..16676
/rpc_family="ERV"
repeat_region /rpc_family="ERV"
16970..17198
/rpc_family="L1"
repeat_region /rpc_family="L1"
17199..17548
/rpc_family="MER1_type"
17575..18261
/rpc_family="L1"
repeat_region /rpc_family="L1"
19409..19531
/rpc_family="L1"
repeat_region /rpc_family="L1"
19540..19659
/rpc_family="Alu"
25486..25822
/note="similar to BST AL928247 (NID:G5664211) w09sh11.x1"
misc_feature /note="similar to BST AL928247 (NID:G5664211) w09sh11.x1"
26074..26366
/rpc_family="Alu"
repeat_region /rpc_family="Alu"
26384..26542
/rpc_family="MIR"
repeat_region /rpc_family="MIR"
26569..27058
/rpc_family="L2"
27140..27647
/rpc_family="L2"
repeat_region /rpc_family="L2"
30824..31150
/rpc_family="MaLR"
repeat_region /rpc_family="MaLR"
32267..32525
/rpc_family="CR1"
repeat_region /rpc_family="CR1"
32568..32770
/rpc_family="MER1_type"
32775..33032
/rpc_family="CR1"
repeat_region /rpc_family="CR1"
33172..33418
/rpc_family="MIR"
repeat_region /rpc_family="MIR"
34930..35230
/rpc_family="Alu"
repeat_region /rpc_family="Alu"
35302..35555
/rpc_family="L2"
35975..36081
/rpc_family="L2"
repeat_region /rpc_family="L2"
36364..36515
/rpc_family="Alu"
repeat_region /rpc_family="Alu"
37266..37340
/rpc_family="L2"
repeat_region /rpc_family="L2"
37400..37440
/rpc_family="L2"
repeat_region /rpc_family="L2"
37615..37715
/rpc_family="MIR"
repeat_region /rpc_family="MIR"
38006..38177
/rpc_family="MER1_type"
repeat_region /rpc_family="MER1_type"
43171..43204
/rpc_family="U2"
repeat_region /rpc_family="U2"
44777..44926
/rpc_family="MIR"
repeat_region /rpc_family="MIR"
45905..46110
/rpc_family="MIR"
```

```
repeat_region 48479..48638
/rpc_family="MER1_type"
repeat_region 48914..49033
/rpc_family="MIR"
repeat_region 49069..49337
/rpc_family="MER1_type"
repeat_region 49340..49621
/rpc_family="Alu"
repeat_region 50377..50498
/rpc_family="MIR"
repeat_region 56393..56545
/rpc_family="MIR"
repeat_region 58109..58420

Alignment Scores:
Pred. No.: 2,24e-89 Length: 168651
Score: 2221.50 Matches: 486
Percent Similarity: 68.10% Conservative: 67
Best Local Similarity: 59.85% Mismatches: 139
Query Match: 45.25% Indels: 131
DB: 9 Gaps: 9

US-09-720-086-5 (1-908) x AC009474 (1-168651)
QY 206 TrpLysArgGluAlaGluLysAlaLysValIleAlaValMetAsnAlaValGlu 225
Db 74844 TGGCCTCATGGGCTAGAGAGAAAGCCGAGGCTCACTGCATTAATGATGAGAGAA
QY 226 AsnGlnAlaSerGlyLysSerGlnLysValGluAlaSerProProAlaValGln 245
Db 74904 AACCTAGGCTCAGAGAGAGCTTCAGAAAGGTGAGAGGCCAGCCCTGCTGTGGCAG
QY 246 ProThrAspProAlaSerProThrValAlaThrThrProGluProValGlyAspAla 265
Db 74964 CCCAGCAACCCACACTCCCCCACTGCTTACCAACCTGAGCCTGTGGGAGCTGAGCT
QY 266 GlyAspLysAsnAlaThrLysAlaAlaAspAspGluProGluTrpGluAspGly 285
Db 75024 GGGAAACAG---TACATCAAGCAGGTGACGATGAACAGATCAAGACAGCTGGGC
QY 286 PheGlyIleGlyGluLeuValTrpGlyLysLeuArgGlyPheSerTrpTrpProGlyArg 305
Db 75081 TTTGGCAATGGGAGAGCTGTGTAGAGAAACTGCAAGGCTTCTCTGTGGTGGCCG---
QY 306 IleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGluGlyThrArgTrpValMet 325
Db 75136 ATTGTCTTGTGTGATGACAGCCCAAGCCGACGCTGAAGCAACCACTGTATC
QY 326 TrpPheGlyAspGlyLysPheSerValCysValGluLysLeuMetProLeuSerSer 345
Db 75196 TGGCTTGGAGA-GGTGAGTCTTCACTGTGTGTGTAGAGAGCTGATCACTGAGCTCC
QY 346 PheCysSerAlaPheHisGlnAlaThr-----TyrAsnLysGlnProMetTyrArgLys 363
Db 75255 TTTTGCAGTGTGTTCACACAGACCAC-CTATAATACACAGACGCCATGATCACCAAA
QY 364 AlaIleTyrGluValIleGlnValAlaSerSerArgAlaGlyLysLeuPheProAlaCys 383
Db 75314 GCCATCTACGAAGTCTGCGCGGTGGCCAGCACTGT-ACAGGGAGCTTCTTCAGGAGC
QY 384 HisAspSerAspGluSerAspSerGlyLysAlaValGluValGlnAsnLysGlnMetIle 403
Db 75373 CATGACAGTGTATGACAGTACACTGCCAAGATACAGAGTGCAGACAGAGATATTC
QY 404 GluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGluProProGluGlu 423
Db 75433 AGATGACCTTTGGGGGGTTCAGCCTTCTGCCCCAAGATGAGAGCCACAGAGAG
QY 424 GluLysAsnProTyrLysGluValTyrThrAspMetTrpValGluProGluAlaAla 443
Db 75493 GAGAAAGATCCACAGAAAGAGTTTATGACACAGTGAAGGGGTGACCTGAGCAACTGCT
QY 444 TyrAlaProProProAlaLysLysProArgLysSerThrThrGluLys----- 460
```

```

Db 75553 TAGGCACCAACCCACAGCCAAAGAGCCCGAAAGACACAGCTAAGAAAGCTCAACATG 75612
Qy 460 -----
Db 75613 GGACCAAGCACAGTAGCTACACCTGTAAATCCACAGCACCTTTGGGAGGCTGAGCGCGCG 75672
Qy 460 -----
Db 75673 ATCACAGCTCAGGAGATAGAGACCTGACATGGTCAGCCCAACATGGTGAACCCCGCTT 75732
Qy 461 -----
Db 75733 CTACTAAAAAGTACCAAAATAGCTGGACATGGTGGCATGTGCTGTAATCCAGACTACTG 75792
Qy 462 -----
Db 75793 GGGAGGCTGAGGCGAGGAATCACTTGAACAGGAGGTCGGAGGTTACAGTGAGCGGAGA 75852
Qy 462 -----
Db 75853 TAGCACCATTCGACTCCAGCCTGGTGACAGAGCAAGACTCTGTTTCAAAAAAATAAAA 75912
Qy 463 -----
Db 75913 AAAAAAGCTCGACATCAAGGAGATTGTTGATGAAC--ACAGAGAGTAGTTAGTGACAA 75970
Qy 478 uValArgGlnLysCysArgAsnIleGluAapIleCysIleSerCysGlySerLeuAsnVa 498
Db 75971 ACTGAGGAAGAGTGGCAGAACATCAAGGACCTTTGTGCTCTTGTGAGAGCCTCAATGT 76030
Qy 498 lThrLeuGluHisProLeuPheIleGlyMetCysGlnAsnCysLysAsnCysPheLe 518
Db 76031 CACTCTGGAACACCCCTCTTCATGCTAGGAATATGACGACATTCGCAAGAACCTGCTTCT 76090
Qy 518 uGluCysAlaTyrGlnTyrAspAspGlyTyrGlnSerTyrCysThrIleCysCysGl 538
Db 76091 GGAATATGATACAGTATGATGAAGAGGCTATCAGTTCGTGTCACCATCTGCTGCTG 76150
Qy 538 yGlyArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCy 558
Db 76151 TGGGCGCAGGCTCATGTGCTG---AACAACCTGCTGCGAGGCTTTTGGCTGGAGTG 76207
Qy 558 sValAspLeuLeuValGlyProGlyAlaAlaGlnAlaAlaIleLysGluAspProTyrAs 578
Db 76208 TATAGACCTCTTGGTGGCCAGAGGTGGCCAGGTGACCATTAAGGAGATCCCTGGNA 76267
Qy 578 nCysTyrMetCysGlyHisLysGlyThrTyrGlyLeuLeuArgArgGluAspTyrPr 598
Db 76268 CTGTACATGTGTGAACACAAAGGCGATCTGTGAGCTGCTGCGAGTGGCGGGATGACTGCC 76327
Qy 598 sSerArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProLysVa 618
Db 76328 CTCTCAGCTCCAGATGTTCTTCCCAATAACCTTAACAGGAATTTGACCCCTCTGGAAGT 76387
Qy 618 lTyrProProValProAlaGluLysArgLys---ProIleArgValLeuSerLeuPheAs 637
Db 76388 CTACTCGCTGTCTCCAGCTGAGAAAGAGAGCCGCCCATCCAGGTGCTGCTCTTTTGA 76447
Qy 637 pGlyIleAlaThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIl 657
Db 76448 CGGGATTGATATGAGGCTCTGCTGCTGTAATGGCTTGGCATTTAGGTAGATCACTACAT 76507
Qy 657 eAlaSerGluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIl 677
Db 76508 GTC--TCAGAGGTGTGTGAGGACGCCATCATTTGGGGCATGCTGTGGCACCGGAGGAT 76566
Qy 677 eMetTyrValGlyAspValArgSerValThrGlnLysHisIleGlnGluTyrGlyProPh 697
Db 76567 CATGCACATCAGGGATACCGGAACCATCATACAGAACACACACAGGAGTGGGGCCCAT 76626
Qy 697 eAspLeuValIleGlyGlySerProCysAsnAspLeuSerIleValAsnProAlaArgly 717

```

```

Db 76627 CTC-----TGCATGACCTCTCCATTGTCAACCTGCCACGA 76664
Qy 717 sGlyLeuTyrGluGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAl 737
Db 76665 AGGACCCCTACAAGGCACTGGCCAGCAATTTCTTTCAGTTTTCTCACCTGTG-CATCATGC 76723
Qy 737 aArgProLysGluGlyAspArgProPhePheThrLeuPheGluAsnValValAlaMe 757
Db 76724 ACAGCCCAAGAGGAGATGATCA-CCCTTCTTGTGGCTCTTGTGAGAGTGTGGTGGCAT 76782
Qy 757 tGlyValSerAspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAs 777
Db 76783 GGGCACTAGTCAAAAGAGAGACATCTCAAAATTTTATTCCAAATTCATGATGATTA 76842
Qy 777 pAlaLysGluValSerAlaAlaHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMe 797
Db 76843 TGCCAGGAAGTGTCACTGTGCACACCACTACTACTACTCTGTGTAACTCTTCCCAGTAA 76902
Qy 797 tAsnArgProLeuAlaSerThrValAsnAspLysLeuGluGlnGluCysLeuGluHi 817
Db 76903 GAGCAGGCCATTTGGCATTCTACTATGATGT-AAGCTGGAGCTGCAGGAGTGTGTGGAGCA 76961
Qy 817 sGlyArgIleAlaLysPheSerLysValArgThrIleThrArgSerAsnSerIleLy 837
Db 76962 TGCAGAAAAAGCCAAAGCTCAGCAAAGCGAGGACCATTACTACAGGTCAAACTCCATAAT 77021
Qy 837 sGlnGlyLysAspGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTyrCy 857
Db 77022 GCAGGGCAAGACAGCGCTCTCCCTGCTTTCATGAATGAGAAG---GACACCTTAGG-TG 77077
Qy 857 sThrGluMetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSe 877
Db 77078 CATTAATTAATGAAAGGCTATTTGGACTCCCTGTCTTATATCTGACATCTCCAACATAAG 77137
Qy 877 rArgLeuAlaArgGlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLe 897
Db 77138 CCATTGGGCAAGCGAGAGACTGTTGCGTGAGTCATGGAGCATAGCAGTCACTTGCCCAATT 77197
Qy 897 upheAlaProLeuLysGluTyrPheAlaCysVal 908
Db 77198 CTTCCCTCCACTGAAGGGGTATTTTGTGTGTG 77231

RESULT 8
AF068627 4135 bp mRNA linear ROD 06-DEC-199
LOCUS Mus musculus DNA cytosine-5 methyltransferase 3B2 (Dmmt3b) mRNA,
DEFINITION alternatively spliced, complete cds.
ACCESSION AF068627
VERSION AF068627.2 GI:6449471
KEYWORDS SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 4135)
AUTHORS Okano,M., Xie,S. and Li,E.
TITLE Cloning and characterization of a family of novel mammalian DNA
JOURNAL Nat. Genet. 19 (3), 219-220 (1998)
MEDLINE 98324766
PUBMED 9662389
REFERENCE 2 (bases 1 to 4135)
AUTHORS Xie,S., Okano,M. and Li,E.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,
Charlesstown, MA 02129, USA
REFERENCE 3 (bases 1 to 4135)
AUTHORS Okano,M., Chijiwa,T., Sasaki,H. and Li,E.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,
Charlesstown, MA 02129, USA
REMARK Sequence update by submitter
COMMENT On Nov 18, 1999 this sequence version replaced gi:3327981.

```







```
Qy 423 uGlu---LysAsnProTyrLysGluValTyrThrAspMetTrpValGluProGluAlaAl 442
Db 1348 GCAACCCAGAGAACAAAGTCCAGAGCCACCAACCAAT-----GACTCTGC 1392
Qy 442 aLaTyrAlaProProProAlaLysLysProArgLysSerThrThrGluLysProly 462
Db 1393 TGCTTCTGAGTCCCCACCACCAAGCCCTCAAGACAAATAGCTATGGGGGAAGACCG 1452
Qy 462 sValLysGluIleIleAspGluArgThrArgGluArgLeuValTyrGluValArgGlnly 482
Db 1453 AGGGGAG-----GATGAGGAGAGCCGAGACGATGCTCTCTGAAGTCCACCAACAA 1503
Qy 482 sCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThrLeuGluHi 502
Db 1504 CAAGGCCATCTGGAACACCGCTGTTTCTGTCTGTGGAAAGAGAACCTGTGTCTCTCCA 1563
Qy 502 sProLeuPheIleGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGluCysAlaTy 522
Db 1564 CCCCCTCTTTGAGGGTGGGCTCTGTCAAGATTGCGGGATCGCTTCTTAGAGCTCTTTA 1623
Qy 522 rGlnTyrAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyGlyArgGluVa 542
Db 1624 CATGTATGATGAGGCGCTATCAGTCTACTGCACCGTGTCTGTGAGGGCGGTGAAC 1683
Qy 542 lLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysValAspLeuLe 562
Db 1684 GTCTGTGTGAGTAAACAAGCTGCTGCAGATGCTTCTGTGTGGAGTGTCTGGAGGTCT 1743
Qy 562 uValGlyProGlyAlaAlaGlnAlaIleLysGluAspProTrpAsnCysTyrMetCy 582
Db 1744 CGTGGCGCAGGACACAGCTGAGGATGCCAAGCTGCAGAACCTCGAGCTGTATATGTG 1803
Qy 582 sGlyHisLysGlyThrTyrGlyLeuLeuArgArgGluAspTrpProSerArgGluGl 602
Db 1804 CTTCCCTCAGCGTGCATGGGTCTCCGACGAGGAGAAAGATTGGAACATCGCGCTGCA 1863
Qy 602 mMetPheAlaAsnAsnHisAsp---GlnGluPheAspProProLysValTyrProPr 621
Db 1864 AGACTTCTTCTACTACTGATCTGACCTGGAAGAAATTTGAGCCACCAAGATTGTACCCAGC 1923
Qy 621 oValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIleAlaTh 641
Db 1924 AATCTCTGACGACCAAGAGGAGCCCATAGAGTCTGTCTGTGTGATGAATGTCAC 1983
Qy 641 rGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSerGluVa 661
Db 1984 GGGGTACTTGGTGTCTCAAGGAGTGGGTATTAAAGTGGAAAGTACATTGCCCTCGAAGT 2043
Qy 661 lCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyrValGl 681
Db 2044 CTGTGCAGAGTCCATCGCTGTGGAACTGTTAAGCATGAAGGCCAGATCAAAATATGTCAA 2103
Qy 681 yAspValArgSerValThrGlnLysHisIleGlnGluTrpGlyProPheAspLeuValIl 701
Db 2104 TGACGTCCGGAATAATCACCAGAAATAATTTGAAGTGGGGCCGCTTCGACTTGGTGTAT 2163
Qy 701 eGlyGlySerProCysAsnAspLeuSerIleValAsnProIleArgLysGlyLeuTyrGl 721
Db 2164 TGGTGAAGCCCATGCAATGATCTCTTAAGCTCAATCTCCCGCCAAAGGTTTATATGA 2223
Qy 721 uGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgProLysGl 741
Db 2224 GGGCACAGGAAGGCTCTTCTCGAGTTTACCACCTTGTCTGATATATACCGCCCAAGGA 2283
Qy 741 uGlyAspArgProPhePheThrLeuPheGluAsnValValAlaMetGlyValSerAs 761
Db 2284 GGGCGCAACCCGTCATCTCTGAGATTGTCGAGAATGTTGTGGCCATGAAGTGAATGA 2343
Qy 761 pLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLysGluVa 781
Db 2344 CAAGAAAGACATCTCAAGATTCTCTGGCATGTAAACCCAGTGTATCGATGTCATCAAGT 2403
```

```
Qy 781 lSerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMetAsnArgProLe 801
Db 2404 GTCTGCTGCTCACAGGCGCCGCTACTTCTGGGTGTAACCTACCCGAATGAACAGGCCCG 2463
Qy 801 uAlaSerThrValAsnAspLysLeuGluLeuGlnGluCysLeuGluHisGlyArgIleAl 821
Db 2464 GATGGCTTCAAGAGATGATAAGCTCGAGCTCGAGGCTGCCTGGAGTTCACTAGGACAGC 2523
Qy 821 aLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGlyLysAs 841
Db 2524 AAAGTTAAAGAAAGTGCAGACAATAACCAAGTCGAACCTCCATCAGACAGGCAAAA 2583
Qy 841 pGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTyrCysThrGluMetGl 861
Db 2584 CCAGCTTTCCCTGTAGTCATGAATGGCAAGGACGACGCTTTTGTGTGCTGACCTGAGT 2643
Qy 861 uArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeuAlaAr 881
Db 2644 AAGATCTTCGGCTTCCTGCTGCTCACTACAGGACGTGTCCAAACATGGCGCGCGCCG 2703
Qy 881 gGlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAlaProLe 901
Db 2704 TCAGAAGCTGTGGCAGGTCCTGGAGTGTACCGGTGTCATCAGACACCTGTTGCCCCCT 2763
Qy 901 uLysGluTyrPheAlaCys 907
Db 2764 GAAGGACTACTTTGCCCTGT 2782
```

RESULT 9

AF151970 4278 bp mRNA linear ROD 08-JUN-2000

LOCUS Mus musculus DNA cytosine-specific methyltransferase isoform 2 (Dnmt3b) mRNA, complete cds.

DEFINITION AF151970

ACCESSION AF151970

VERSION AF151970.1 GI:8347119

KEYWORDS Mus musculus.

SOURCE Mus musculus.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 4278)

AUTHORS Shen, Y., Chen, Y.T., Zhu, M., Luo, Y.J., Zhu, N., Xu, S.C., Wu, G.Y. and

TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing

JOURNAL isoforms in mouse embryonic tissue

REFERENCE 2 (bases 1 to 4278)

AUTHORS Yin, B., Chen, Y.T., Zhu, M., Luo, Y.J., Zhu, N., Xu, S.C., Wu, G.Y. and

TITLE Direct Submission

JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular

REFERENCE Biology, Institution of Basic Medical Sciences, Chinese Academy of

AUTHORS Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,

P.R.China

FEATURES Location/Qualifiers

1..4278

/organism="Mus musculus"

/strain="KM"

/db\_xref="taxon:10090"

/dev\_stage="8-9 day old embryo"

1..4278

/gene="Dnmt3b"

412..2931

/gene="Dnmt3b"

/function="methylates cytosine in DNA"

/note="alternatively spliced"

/codon\_start=1

/product="DNA cytosine-specific methyltransferase isoform 2"

/protein\_id="AAF74516.1"

/db\_xref="GI:8347120"

/translation="MKGDSRLNNEEGASGYEECIIVNGNFSQSDSDTKDAPPPVLE

AICTPVCPTETRRSSRLSKREVSLLNTQDMTGDGRDDEVDGNGSDILMPK



```
Qy 542 lLeuMetCysGlyAAsnAsnAAsnCysCysArgCysPheCysValGluCysValAspLeuLeu 562
Db 1827 GCTGCTGTGCAGTAACACACAGAGCTGCTGCAGATGCTTCTGTGGAGTGTCTGGAGGTGCT 1886

Qy 562 uValGlyProGlyAlaGlnAlaAlaIleLysGluAspProTrpAsnCysValMetCys 582
Db 1887 GGTGGCGCAGGACAGCTGAGGATGCACAGCTGCAGAACCTTGGAGCTGCTATATGTG 1946

Qy 582 sGlyHisGlyThrTyrGlyLeuLeuArgArgGluAspTrpProSerArgLeuG1 602
Db 1947 CTTCCCTCAGCGCTGCATGGGTCTCTCCGACGAGGAAAGATTGGACATCGCGCTGCA 2006

Qy 602 nMetPhePheAlaAsnHisAsp---GlnGluPheAspProProLysValTyrProPr 621
Db 2007 AGACTTCTTCACTACTGATCTGACCTGGAAGATTGAGCCACCCAGATTGATCCAC 2066

Qy 621 oValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIleAlaTh 641
Db 2067 AATTCTGTCAGCCAAAGAGGCCATATGAGTCTGCTCTGTGTGATGGAATTGCAAC 2126

Qy 641 rGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSerGluVa 661
Db 2127 GGGGTACTTGGTCTCAAGGAGTTGGGTATTAAAGTGAAGATACATTGCTCCGAAGT 2186

Qy 661 lCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyrValGl 681
Db 2187 CTGTGCAGAGTTCATCGCTGTGGAACTGTTAAGCATGAAGCCGAGATCAATATGTCAA 2246

Qy 681 yAspValArgSerValThrGlnLysHisIleGlnGluTrpGlyProPheAspLeuVal1 701
Db 2247 TGACGTCCGGAATAACCAAGAAATAATTGAAGAGTGGGGCCGCTTTCGACTTGGTGA 2306

Qy 701 eGlyLysSerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeuTyrGl 721
Db 2307 TGGTGAAGCCCATGCAATGATCTCTAAGCTCAATCTCCCGCAAGGTTTATATGA 2366

Qy 721 uGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgProLysGl 741
Db 2367 GGGCACAGGAAGGCTCTTCTCGAGTTTATACCACTTGTGTAATATATACCCGCCCAAGGA 2426

Qy 741 uGlyAspAspArgProPhePheTrpLeuPheGluAsnValAlaAlaMetGlyValSerAs 761
Db 2427 GGGCCACACACCGTCCATTCTTCTGGATGTTGAGATGTTGTGCCATGAAGTGAATGA 2486

Qy 761 pLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLysGluVa 781
Db 2487 CAAGAAAGACATCTCAAGATTCCTGCGCATGTAAACCCAGTGTATGATGATGATGATGAT 2546

Qy 781 lSerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMetAsnArgProLe 801
Db 2547 GTCTGCTGTCTACAGGCCCGGTACTTCTGGGTAACTTACCCGGAATGAACAGGCCCGT 2606

Qy 801 uAlaSerThrValAsnAspLysLeuGluGlnGluCysLeuGluHisGlyArgIleAl 821
Db 2607 GATGCTCTCAAGAAATGATAAGCTCGAGCTGCAGAGTGTGCTGAGTTCAGTAGACAGC 2666

Qy 821 aLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGlyLysAs 841
Db 2667 AAAGTTAAAGAAAGTGCAGACATAATACCAACCAAGTCAAGTCACTACACAGGGCAAAA 2726

Qy 841 pGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTrpCysThrGluMetGl 861
Db 2727 CCAGCTTCTTCCCTGTGATCATGAATGGCAAGGACGCTTGTGTGTCATGATGATGATGAT 2786

Qy 861 uArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeuAlaAr 881
Db 2787 AAGATCTTCCGCTTCCCTGCTCACTACACGAGCTGTCCAAACATGGCGCGCGCGCG 2846

Qy 881 gGlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAlaProLe 901
Db 2847 TCAGAGCTGCTGGGAGGCTGCTGGAGTGTACCGGTCTATCAGACACCTGTTTGGCCCC 2906

Qy 901 uLysGluTyrPheAlaCys 907
```

```
||||:|||||||||
Db 2907 GAAGGACTACTTTGCTGT 2925

RESULT 10
AF068626 4195 bp mRNA linear ROD 06-DEC-1999
LOCUS Mus musculus DNA cytosine-5 methyltransferase 3B1 (Dnmt3b) mRNA,
DEFINITION alternatively spliced, complete cds.
ACCESSION AF068626
VERSION AF068626.2 GI:6449469
KEYWORDS SOURCE
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 4195)
Okano,M., Xie,S. and Li,E.
Cloning and characterization of a family of novel mammalian DNA
(cytosine-5) methyltransferases
Nat. Genet. 19 (3), 219-220 (1998)
JOURNAL 98324766
MEDLINE 9662389
PUBMED
REFERENCE 2 (bases 1 to 4195)
Xie,S., Okano,M. and Li,E.
Direct Submission
Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,
Charlestown, MA 02129, USA
JOURNAL 3 (bases 1 to 4195)
Okano,M., Chijiwa,T., Sasaki,H. and Li,E.
Direct Submission
Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,
Charlestown, MA 02129, USA
REMARK Sequence update by submitter
COMMENT On Nov 18, 1999 this sequence version replaced gi:3327979.
FEATURES
Location/Qualifiers
1..4195
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="2"
/map="84.0 CM"
1..4195
/gene="Dnmt3b"
1..4195
/gene="Dnmt3b"
/notes="Dnmt3b"
/notes="similar to EST sequences deposited in GenBank
Accession Numbers AA116694, AA119579, AA177277, AA210568,
AA407106, and AA575617"
269..2848
/gene="Dnmt3b"
/function="de novo DNA methylation"
/notes="alternatively spliced product; contains Cys-rich
region; C-terminal region is similar to corresponding
region of DNA cytosine-5 methyltransferase 3A"
/codon_start=1
/product="DNA cytosine-5 methyltransferase 3B1"
/db_xref="GI:6449470"
/translation="MKGDSRHLNEEGASGVVECIIVNGNFSDOSSDTKADSPVPVLE
ALCTPEVCTPETGRSSRLSKREVSSLLNYTODMTGDRDGDDEVDGNGSDILMPK
LPTREKDTSTRESPAVTRTHSNGTSLERQASPRITRGQGHVVEYVPEPATR
SRRRASSTSPWSPASVDPMEEVTPKSVSTPSVLSQDQDGMITTVQDAESRD
GSTETFDKKEFGIGLVKIGFWSWPAWVSWKATSKRQAMPGRWQVFGDKF
SEISAKLVAGLFQHFNLATKLVSYRKAMVHTLEKARVAKGTFTSSSSLEL
QULKMLEAHGGFKPTGIEGLKPNKQPVVNSKVRSDNSLRPRRKRRTTN
DAAASESPPKLKTNSYGGKDEDESERMASVETNNKGNLLECLSCGKNPVS
PHLPFEGGLQCSDRDLFLFYMDGQSYCTVCCGRELILCSNTSCCFCVCEP
LVLVAGTAEDAQLQEPWCMCLPORCHGLVRRKDMNRLQDFTTDPLEEFEP
PKLYPAIPAAKRRPIRVLSLDGIATGVLVKEIGIKVEKVIASEVCAESIAGTVKH
EQGIKVNDRVKTNNIEEWGPFDLVIGGSPCNLSNVNPARGLYSGTGLPPEFY
HLNLTVPKXGDNRPFFWPFENVAMKVNKKDLSRLACNPMIDAIVKSAAHARY
FWGNLPGMNRPMWASKNDKLELDQCLFSTRATKLKVTITTKNSIRKGNKPLPVV
MNGKDDVLWCTELERIFGFFAHTYDVSNMGRGARQKLLGRSWSVPVIRHLFAPLKYDF
ACE"
```





QY 38 oSerAlaThrAlaArgIySeValGIyArgProGIyValArgIySerHisPro---ProVa 57  
 Db 285 CGACAGACCCCGCTGAGGCTTGTGCGACCTTGGAACCTCAGATATATACCTTTCAGCA 344  
 QY 57 IGIuSerSerAspThrProIySAspProAlaValThrThrIySerGIuProMetAlaGI 77  
 Db 345 CGGGGATCTCTCCCTCCCATCCATAGTGCCTGGGACCAATCCAGGCGCTTCTTCA 404  
 QY 77 nAspSerGIyProSerAsp----- 83  
 Db 405 GGAACAATATGAAGGAGACAGACACTGTGAATGAAGAAGCGGTCCAGCGGTATGA 464  
 QY 84 -----LeuLeuProAsnGIyAspLeuGIuIySAspSerGIuProGIuProGIuGIuGI 101  
 Db 465 GGAGTGCATTTATCTGTTAATGGAACTTCAGTACCACTCTCAGACACGAAGATCTCC 524  
 QY 101 ySerProAlaAlaGIyGIuIySGLyGIyAlaProAlaGIuGIyGIuGIyThrGIuThrPr 121  
 Db 525 CTCACCC-----CC 533  
 QY 121 oProGIuAlaSerArgAlaValaGIuAsnGIyCysCysValThrIySGLyGIyArgGIyAl 141  
 Db 534 AGCTTGGAGGCAATCTGCACAGHCCAGCTTCACACACAGACAGAGCGCCGACGCTC 593  
 QY 141 aSerAla----- 143  
 Db 594 AGCTCCCGCTGCTTAAAGAGGAGTCTCCAGCCTTGTGAATTACACGACGACATGAC 653  
 QY 144 -GIyGIuGIy----- 146  
 Db 654 AGGAGATGGAACAGAGATGATGATGATGATGGGAATGGCTGTGATTCTTAATGCC 713  
 QY 147 -----LysGIuGIuIySGLInThrAsnIleGIuSerMetIySGLyGIySerAr 163  
 Db 714 AAAGCTCACCCGTGAGACCAAGACACACAGACGCGCTCT-----GAAAGCCCGGC 764  
 QY 163 gGIyArgIyLeuArgGIyGIyLeuGIyTrpGIuSerSerLeuArgGIuInArgProMetProAr 183  
 Db 765 TGCTCCGAACCCGACATAGCAATGGGACCTCCAGCTTGGAGAGGCAAGAGCCTCCCCAG 824  
 QY 183 gLeuThrPheGIuAlaGIyAspProIyTrpIleSerIySArgIySAspGIuTrpLe 203  
 Db 825 AATCACCCGAGGTGCGCAGAGGCCGCCACATGT-----CAGAGATACC 869  
 QY 203 uAlaArgTrpIySArgIuAlaGIuIySGLySGLyAlaIySValIleAlaValMetAsnAlaVa 223  
 Db 870 TGTGGAGTTTCCGCTACCAAGTCTCGAGACGTGACGATCTGCTTACAGCAAGCAGCC 929  
 QY 223 IGIuGIuAsnGIuAlaSerGIyGIuSerGIuIySValGIuGIuIyLaserProProAlaVa 243  
 Db 930 ATGGTCAATCCCTGCGCAGCGTGCACCTC-----ATGGAAGAGTGCACCTTAAGAGCGT 983  
 QY 243 IGIuGIuProThrAspProAlaSerProThrValAlaThrThrProGIuProValGIyGI 263  
 Db 984 CAGTACCCCA-----TCAGTTGACTTGACCCAGATGAGATCAGAGGCTAT 1031  
 QY 263 yAspAlaGIyAspIySAsnAlaThrIySAlaAlaAspAspGIuProGIuTrpGIyGIuAspGI 283  
 Db 1032 GGATACACACAGAGTGCAGAGACATATATGAGACAGCAGCAGAGTATCAGATGA 1091  
 QY 283 yArgGIyPheGIyIleGIyGIuLeuValIyTrpGIyIySLeuArgGIyPheSerTrpTrp 303  
 Db 1092 TAAAGAGTTTGGAAATAGTGAACCTCGTGTGGGAAGAATCAAGGCTTCTCTGTGGGCC 1151  
 QY 303 oGIyAspGIyIleValSerTrpTrpMetThrGIyArgSerArgAlaAlaGIuGIyThrArgTr 323  
 Db 1152 TGGCATGTGTGTCTCTGGAAGCAGCTCCAGAGCAGACAGGCCATGCGCGAGATGGCTG 1211  
 QY 323 pValMetTrpPheGIyAspGIyIySPheserValValCysValGIuIySLeuMetProLe 343  
 Db 1212 GGTACAGTGTGTGTGTGATGGCAAGTTTCTGAGATCTGTGCTACAAACTGTGTGCTCT 1271

QY 343 uSerSerPheCysSerAlaPheHisGIuAlaThrTyraSLeuIySGLInProMetTyArgIy 363  
 Db 1272 GGGGCTGTTCAGCCAGCACTTATCTGCTTACCTTCAATTAAGCTGTTCTTATAGGAA 1311  
 QY 363 sAlaIleTyGIuValLeuGIuValAlaSerSerArgAlaGIyIySLeuPheProAlaIy 383  
 Db 1332 GGCTATGTACCACTCTGGAGAAAGCAGGCTTCAGCTGGCAAGACCTTC----- 1393  
 QY 383 sHisAspSerAspGIuSerAspSerGIySAlaValaGIuValaGIuIySLeuIyMetI 403  
 Db 1384 -----TCCAGAGTCTCTGGAGAGTCACTGGAGGACAGCTGAAGCCATGCT 1430  
 QY 403 eGIuTrpAlaLeuGIyGIyPheGIuInProSerGIyProIySGLyLeuGIuPro----- 420  
 Db 1431 GGAGTGGCCCAAGTGGCTTCAAGCTTACGGATCGAGGCGCTCAAAACCAACAGAA 1490  
 QY 421 -----ProGI 422  
 Db 1491 GCACACAGTGTATTAAGTCCGAGGTGCGTTCAGACAGTAGGAACCTTAAGAACCCAG 1550  
 QY 422 uGIuGIuIySAsnProIyTrpIySGLuValTyThrAspMetTrpValGIuProGIuAlaI 442  
 Db 1551 GAGACCGGAGAACAAAGTCCGAAGACGACACCAAT-----GACTCTGC 1595  
 QY 442 aAlaTyAlaProProProProAlaIySValSAspArgIySerThrGIuIySProIy 462  
 Db 1596 TGCTTGTAGTCCCCCCCCCAAGCGCTTCAGACAAATAGCTATGGCGGAGAGACCG 1655  
 QY 462 sValIySGLuIleIleAspGIuArgTrpArgGIuArgLeuValTyTrpGIuValaArgGIuIy 482  
 Db 1656 AGGGAG-----GATGAGAGAGCCGAGAACGATGCTTCTTAAGTACCAACAA 1706  
 QY 482 sCysArgAsnIleGIuAspIleCysIleSerCysGIySerLeuAsnValThrLeuGIuI 502  
 Db 1707 CAAGGGCAATCTGGAAGACCGCTGTGTCTGTGTGGAAGAAGAACCTGTGTCTTCCA 1766  
 QY 502 sProLeuPheIleGIyGIyMetCysGIuAsnCysIySAsnCysPheLeuGIuIyCysAlaTy 522  
 Db 1767 CCCCCTTTGAGAGTGGTGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1826  
 QY 522 rGIuTrpAspAspAspGIyTyGIuInSerTyCysThrIleCysCysGIyGIyArgGIuVa 542  
 Db 1827 CATGTATGATGAGAGCGGTATCTAGTCTTACTGCACCGTGTGTGTGTGTGTGTGTGT 1886  
 QY 542 ILeuMetCysGIyIySAsnAsnAsnCysCysArgCysPheCysValaGIuIyCysVala 562  
 Db 1887 GGT 1946  
 QY 562 uValGIyProGIuAlaAlaGIuAlaAlaIleIySGLuSAspProTrpAsnCysTyMetCy 582  
 Db 1947 GGTGGGCGCAGGACAGCTGAGATGCGCAGAGCTGCAGAGAACCTTGAGCTGTATATGTG 2006  
 QY 582 sGIuHisIySGLyThrTyGIyLeuLeuArgArgGIuAspTrpProSerArgLeuGI 602  
 Db 2007 CTTCCCTCAGCGGCTCCATGGGTCTCCGACGCGAGAAAGATTTGGAACATGGCGCTGCA 2066  
 QY 602 nMetPhePheAlaAsnAsnHisAsp-----GIuGIuPheAspProProIySValTyTrpPr 621  
 Db 2067 AGACTTCTTACTTACTGATCTGACCTGGAAGAAATTTGAGCCCAAGTTGATACCAGC 2126  
 QY 621 oValProAlaGIuIySArgIySProIleArgValLeuSerLeuPheAspGIyIleAlaTh 641  
 Db 2127 AATTCGTGACGCAAAAGAGGCGCATTAAGTCTGTCTGTCTGTATGAAATGCAAC 2186  
 QY 641 rGIyLeuLeuValLeuIySAspLeuGIyIleGIuValaAspArgTyTrpIleAlaSerGIuVa 661  
 Db 2187 GGGGTACTTGGTGTCAAGAGAGTTGGGTATTAAGTGAAGAAAGTCAATTGCTCCGAGT 2246  
 QY 661 IySGLuAspSerIleThrValaGIyMetValaArgHisGIuGIyIySLeuTyTrpValGI 681  
 Db 2247 CTGTGCAGAGTCCATCGCTGTGGAAGTGTAAAGCATGAAGGCGCATCAATATGTCAA 2306  
 QY 681 yAspValArgSerValThrGIuIySHisIleGIuGIuTrpGIyProPheAspLeuValI 701



```
Db 2307 TGACGTCGCGAAATATCAACCAAGAAATATTTGAAGAGTGGGGCCGCTTCGACTTGATGAT 2366
Qy 701 eGlyGlySerProCysAsnAspLeuSerIleValAsnProAlaAAGLysGlyLeuTyrGl 721
Db 2367 TGTGGAAGCCCATGCAATGATCTCTTAACGTCATCTGCTCCCAAGAGTTATATGA 2426
Qy 721 uGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaAArgProLysGl 741
Db 2427 GGGCACAGGAAGGCTCTTCTCGAGTTTACCACTTGTGTAATATACCCGCCCAAGGA 2486
Qy 741 uGlyAspAspArgProPhePheTyrLeuPheGluAsnValAlaMetGlyValSerAs 761
Db 2487 GGGCGAACCGCTCTTCTCGAGTGTTCGAGATGTTGTGGCATGAAGTGAATGA 2546
Qy 761 pLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLysGluVa 781
Db 2547 CAAGAAAGACATCTCAAGATTCCTGGCATGTAAACCCAGTGTATGATCGATGCCATCAAGT 2606
Qy 781 lSerAlaAlaHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMetAsnArgProLe 801
Db 2607 GTCTGCTGCTCACAGGCGCCGCTACTTCTGGGGTAACTACCCGGAATGAACAGGCCCT 2666
Qy 801 uLaserThrValAsnAspLysLeuGluLeuGlnCysLeuGluHisGlyArgIleAl 821
Db 2667 GATGGTCTCAAGAAATGATAGCTCGAGCTCGAGACTGCCGTGGAGTTTCAGTAGGACAGC 2726
Qy 821 aLysPheSerLysValArgThrIleThrArgSerAsnSerIleLysGlnGlyLysAs 841
Db 2727 AAGTTTAAAGAAAGTGCAGCAATAACCAACCAAGTCGAATCCATCAGACAGGGCAAAA 2786
Qy 841 pGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTyrCysThrGluMetGl 861
Db 2787 CCAGCTTTTCCCTGTAGTCATGAATGCAAGACAGCAGCTTTGTGGTGACCTGAGCTCGA 2846
Qy 861 uArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeuAlaAr 881
Db 2847 AAGGATCTTCGGCTTCCTGCTCACTACACGACGTGTCCAAACATGGCGCGCGCCG 2906
Qy 881 gGlnArgLeuLeuGlyArgSerTyrPheValProValIleArgHisLeuPheAlaProLe 901
Db 2907 TCAGAGCTCTGGGCAAGTCTCTGAGTGTACCGGTTCATCAGACACCTGTTTGGCCCTT 2966
Qy 901 uLysGluTyrPheAlaCys 907
Db 2967 GAAGGACTACTTTGCTGT 2985

RESULT 12
LOCUS AF151974 4163 bp mRNA linear ROD 08-JUN-2000
DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 6
ACCESSION AF151974
VERSION AF151974.1 GI:8347130
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE 1 (bases 1 to 4163)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
Shen,Y.
Cloning of full-length Dnmt3b cDNA and its alternative splicing
isoforms in mouse embryonic tissue
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4163)
AUTHORS Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
Shen,Y.
Direct Submission
TITLE Submitted (17-MAY-1999) Department of Biochemistry and Molecular
JOURNAL Biology, Institution of Basic Medical Sciences, Chinese Academy of
Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,
P.R.China
```

```
FEATURES
source
Location/Qualifiers
1..4163
/organism="Mus musculus"
/strain="KM"
/db_xref="taxon:10090"
/dev_stage="8-9 day old embryo"
1..4163
/gene="Dnmt3b"
CDS
297..2816
/gene="Dnmt3b"
/function="methylates cytosine in DNA"
/notes="alternatively spliced"
/codon_start=1
/product="DNA cytosine-specific methyltransferase isoform 6"
/protein_id="AAF74520.1"
/db_xref="GI:8347131"
/translation="MKGDSRLHNEEGASGYEECIIVNGNFSQSDTKDAPSPVIT
AICTPEVTPETRRSSRLSKREVSLNNTQMDTGDGDRDDVDGNGSDILMPK
LRTETKDTTRTSSEPAVTRHSNGTSSLERQASPRITRQGRHHVQYVFPFATR
SRRRASSASATPWSPASVDMEETPKSVTPSDLSQDQDQEGMDTQVDAEITY
GDSTEYODDXEFGIGDLVWGKIKGFSWMPVMSWKATSKROAMPKMRWQVDFDGF
SEISADKLVALGLPQSHFNLTNKLVSRYKAMYHTLEKARVAGTKFSSPGESLE
OLKMLEWAGGFKPTGIEGLKPNKKOPENKSRRTTNDNAASESPPKRLKTSYGG
KDRGEDESRERMASEVTNNKLNLEDRLCSGCKNPNVSPHPLFEGGLCQSCRFLLE
FYMIDEGYSYCTVCCGEBLLCSNTSCCCFCVECLVGLVGADEADAKQEPWS
CYMCLPORCHVLRKMDNMRLOQDFTTDPLEBEPFKLPYPAIAKRRPIRVLSL
FDGIATGLVLKELGIVKYEYIASEVCAESIAGTVKHEGQIKYVNDVRKIRKNIIE
WGFDFLVIGGSPCNDLGNVPARKLYEGTGLFFEFYHLLNLTGRKPDGNDRPFWMF
ENVVAMKVNDDKDISRLACNPVMDIAIKVSAARHARYFWGNLPGNRPVMAKNDKL
ELQDCLFESRTAKLVQTIITKNSIRQGNOLFPVVMNGKDDVLWCTELERIFGFP
AHTDVSNMGRGARQKLGRSVPVIRLHAPLAKDYFACE"
BASE COUNT 1032 a 1074 c 1128 g 929 t
ORIGIN
Alignment Scores:
Pred. No.: 3,37e-90 Length: 4163
Score: 2203.50 Matches: 448
Percent Similarity: 61.91% Conservative: 134
Best Local Similarity: 47.66% Mismatches: 236
Query Match: 44.89% Indels: 123
DB: 10 Gaps: 15
US-09-720-086-5 (1-908) x AF151974 (1-4163)
Qy 44 LysValGlyArgProGlyArgLysArgLys----- 53
Db 130 CGCACGGCGCTCCAGTCCGCGCGCGGGGTTAAGCGGCCCAAGTAAACGTAGCGCAG 189
Qy 54 -----HisProProValGluSerSerAspThrProLysAspProAlaValThrLys 71
Db 190 CGATCGGCGCGGAGATTCCGGAACCCGACACTCCGCGCGCGCGCGCGCGCGCGC 249
Qy 72 -----SerGlnProMetAlaGlnAspSerGlyProSerAsp 83
Db 250 GCGCGATCGCGCGCGCGCTTACAGCCAGCCTCA-CGACAGGAAACAATGAAGGAGAC 308
Qy 84 -----LeuLeuProAsn 87
Db 309 AGCAGACATCTGAATGAAGAGAGGTGCGCAGCGGTATGAGGAGTGCAATTCGTTAAT 368
Qy 88 GlyAspLeuGlyLysArgSerGluProGlnProGluGluGlySerProAlaAlaGlyCln 107
Db 369 GGGAACTTCAGTGACCTCTCCAGACGAGAGGATGCTCCCTCACCC----- 416
Qy 108 LysGlyGlyAlaProAlaGluGlyGluGlyThrGluThrProProGluAlaSerArgAla 127
Db 417 -----CCAGTCTTGGAGGCAATCTGC 437
Qy 128 ValGluAsnGlyCysValThrLysGluGlyArgGlyAlaSerAla----- 143
Db 438 ACAGAGCCAGTCTGCACACAGGAGACAGGCGCGCGCGCGCGCGCGCGCGCGCTCTAAG 497
```

QY 144 -----glyglu1y----- 146  
Db 498 AGGAGGCTCTCCAGCTTCTGAATTACACGACGACATGACAGAGATGACAGACAGAT 557  
QY 147 -----lysglu1n 149  
Db 558 GATGAAGTAGATGATGGGAATGGCTGTGATATTCTAATGCCAAAGCTCACCCGTGAGACC 617  
QY 150 Lysgl1ThrAsn1IleGluSerMetLysMetGluGlySerArgGlyArgLeuArgGly 169  
Db 618 AAGGACACCGAGGACCGGCTCT-----GAAAGCCCGGCTGTCGAAACCCGACATAGC 668  
QY 170 LeuGly1TrpGluSerSerLeuArgGlnArgProMetProArgLeuThrPheGln1aGly 189  
Db 669 AATGGACCTCCAGCTTGGAGAGGCAAGAGCTCCCGCAATCACCCGAGGTGGCAG 728  
QY 190 AspProTy1r1IleSerLysArgLysArgSpGluTrpLeu1aArgTrpLysArgGlu 209  
Db 729 GGCCGCCACCATGTG-----CAGAGTACCTGCTGGAGTTTCCGGCTACC 773  
QY 210 AlaGluLysAlaLysAlaValIleAlaValMetAsnAlaValGluGluAsnGln1aSer 229  
Db 774 AGGTCTCGAGACGTCGACGACATGCTTTCAGACAGACGCCATGTATCCCTGCGCAGC 833  
QY 230 GlyGluSerGlnLysValGluGluAlaSerProProAlaValGlnGlnProThrAspPro 249  
Db 834 CTCGACTTC-----ATGGAAGAAGTACACCTTAAGAGCGTCACTACCCA----- 878  
QY 250 AlaSerProThrValAlaThrThrProGluProValGlyGlyAspAlaGlyAspLysAsn 269  
Db 879 ---TCAGTTGACTTGACGACGAGATGAGATCAGAGGATATGATATCACACAGGTGGAT 935  
QY 270 AlaThrLysAlaAlaAspAspGluProGluTrpGluAspGlyArgGlyPheGly1IleGly 289  
Db 936 GCAGAGAGCATATATGAGACACAGACAGATATCAGATGATAAAGCTTTCGAATAGGT 995  
QY 290 GluLeuVal1rTrpGlyLysLeuArgGlyPheSerTrpTrpProGlyArgGlyValSerTrp 309  
Db 996 GACCTCGTGGGGAAGATCAAGGCGCTTCTCCTGCTGCTGCGCATGCTGCTGCTGCTG 1055  
QY 310 TrpMetThrGlyArgSerArgAlaAlaGluGlyThrArgTrpValMetTrpPheGlyAsp 339  
Db 1056 AAGACCACTTCAGCGACGACAGGACATGCCGGAATGCCGCTGGGATCACTGTTGTGAT 1115  
QY 330 GlyLysPheSerValValCysValGluLysLeuMetProLeuSerSerPheCysSerAla 349  
Db 1116 GGCAAGTTTCTGAGATCTCTGCTGACAACTGTGGCTCTGGGCGCTTTCAGCCAGCAC 1175  
QY 350 PheH1Gln1aThrThrAsnLysGlnProMetTy1rArgLysAla1aIleTy1rGluValLeu 369  
Db 1176 TTTAATCTGGCTACTTCAATAGCTGTTCTTATAGGAAGGCGCATGTACACACTGTG 1235  
QY 370 GlnValAlaSerSerArgAlaGlyLysLeuPheProAlaCysHisAspSerAspGluSer 389  
Db 1236 GAGAAAGCCAGGGTTCCAGCTGCGCAAGACTTC-----TCCAGC 1274  
QY 390 AspSerGlyLysAlaValGluValGlnAsnLysGlnMetIleGluTrpAlaLeuGlyGly 409  
Db 1275 AGTCTCGAGAGTCACTGAGGACCACTGAAGCCCATGTGAGTGGGCCCAAGGTGGC 1334  
QY 410 PheGlnProSerGlyProLysGlyLeuGluProProGluGluGlu---LysAsnProTy1r 428  
Db 1335 TTCAGGCTACTGGAGTCAGAGGCGCTCAAAACCAAGAACAGACCCGAGAACAAAGT 1394  
QY 429 LysGluValTy1rThrAspMetTrpValGluProGluAlaAlaAlaTy1rAlaProProPro 448  
Db 1395 CGAAGACGCAACCAAT-----GACTCTGCTGCTTCTGAGTCCCCCCA 1439  
QY 449 ProAlaLysLysProArgLysSerThrThrGluLysProLysValLysGlu1IleLeuAsp 468  
Db 1440 CCCAAGGCGCTCAAGACAAATACCTATGCGGGAAGAGCCGAGCGGAG-----GAT 1490  
QY 469 GluArgThrArgGluArgLeuValTy1rGluValAlaArgGlnLysCysArgAsn1IleGluAsp 488

Db 1491 GAGGAGAGCCGAGAACGATGGCTTCTGAAGTCAACCAACACAGGSCAATCTGGAAGAC 1550  
QY 489 IleCys1IleSerCysGlySerLeuAsnValThrLeuGlnHisProLeuPhe1IleGlyGly 508  
Db 1551 CGCTGTTTGTCTGTGGAAGAAGAACCCCTGTCTCTCCACCCCTTTGAGGCTGGG 1610  
QY 509 MetCysGlnAsnCysLysAsnCysPheLeuGluCysAlaTy1rGlnTy1rAspAspAspGly 528  
Db 1611 CTCTGTCAAGAGTTCGCCGGATGCTTCTCTAGAGCTTCTTACATATATGATGAGACGGC 1670  
QY 529 Ty1rGlnSerTy1rCysThr1IleCysCysGlyValArgGluValLeuMetCysGlyAsnAsn 548  
Db 1671 TATCACTTCTACTCACCGGTGTGCTGTGAGGGCCGTGAAGTGTCTGTGTGAGTAAACACA 1730  
QY 549 AsnCysCysArgCysPheCysValGluCysValAspLeuLeuValGlyProGlyValAla 568  
Db 1731 AGCTGCTGACAGATGCTTCTGTGTGAGTGTCTGAGAGTCTGTGTGGCGGACGACAGCT 1790  
QY 569 GlnAlaAla1IleLysGluAspProTrpAsnCysTy1rMetCysGlyHisLysGlyTy1rTy1r 588  
Db 1791 GAGGATGCCAAGCTGCAGAGAACCTTGAGCTGTATATGTGCTCCCTCAGCCGCTGCCAT 1850  
QY 589 GlyLeuLeuArgArgArgGluAspTrpProSerArgLeuGlnMetPhePheAlaAsnAsn 608  
Db 1851 GGGGTCTCTCGACCCGACGAGAAAGATTTGGAACATGCGCTCGCAAGACTTCTTACTACTGAT 1910  
QY 609 HisAsp--GlnGluPheAspProProLysValTy1rProProValProAlaGluLysArg 627  
Db 1911 CCGACCTGGAAGAATTTGAGGACCAACCAAGTGTATCCAGCAATTCCTGCGACCCAAAAG 1970  
QY 628 LysPro1IleArgValLeuSerLeuPheAspGly1IleAlaThrGlyLeuLeuValLeuLys 647  
Db 1971 AGGCCATTTAGAGTCTGTCTGTGTTGATGGAATTCGACGCGGTACTTGTGCTCAAG 2030  
QY 648 AspLeuGly1IleGlnValAspArgTy1rIleAlaSerGluValCysGluAspSer1IleThr 667  
Db 2031 GAGTTGGGTATTAAAGTGAAGAAAGTACATTGCCCGAAGTCTGTGAGAGTCCATCCGT 2090  
QY 668 ValGlyMetValArgHisGlnGlyLysIleMetTy1rValGlyAspValArgSerValThr 687  
Db 2091 GTGGGAAGCTGTAAAGATGAGAGCGCACAGATCAATATATGTCATGACGTCGGAAAAATCAC 2150  
QY 688 GlnLysHis1IleGlnGluTrpGlyProPheAspLeuVal1IleGlyGlySerProCysAsn 707  
Db 2151 AAGAAAAATATTGAAGATGGGGCCGCTTGCATTTGATGAGGGCACAGGAAGCTCTTC 2210  
QY 708 AspLeuSer1IleValAsnProAlaArgLysGlyLeuTy1rGluGlyThrGlyArgLeuPhe 727  
Db 2211 GATCTCTTAACGTCAATCCCTGCCCAAGGTTTATATGAGGGCACAGGAAGCTCTTC 2270  
QY 728 PheGluPheTy1rArgLeuLeuHisAspAlaArgProLysGluGlyAspAspArgProPhe 747  
Db 2271 TTCGAGTTTATACCACTGCTGATATATACCCGCCCAAGGAGGCGCAACCGTCCATTTC 2330  
QY 748 PheTrpLeuPheGluAsnValAlaAlaMetGlyValSerAspLysArgAsp1IleSerArg 767  
Db 2331 TTCTGGATGTTGGAAGATGTTGTGGCCATGAAGTGAATGACAAAGAACATCTCCACA 2390  
QY 768 PheLeuGluSerAsnProValMet1IleAspAlaLysGluValSerAlaHisArgAla 787  
Db 2391 TTCCTGCGCATGTAAACCGATGATGATGATGCCATCAAGGTGTCTGTGCTCACAGGGCC 2450  
QY 788 ArgTy1rPheTrpGlyAsnLeuProGlyMetAsnArgProLeuAlaSerThrValAsnAsp 807  
Db 2451 CGTACTTCTGGGGTAACTTACCCGGAATGAACAGGCCCGTATGCTTCAAGATGAT 2510  
QY 808 LysLeuGluLeuGlnGluCysLeuGluHisGlyValArg1IleAlaLysPheSerLysValArg 827  
Db 2511 AAGCTGAGCTGACAGAGTCTGTGAGTTCAGTAGAGACAGCAAGGTTAAAGAAAGTGCAG 2570  
QY 828 Thr1IleThrArgSerAsnSer1IleLysGlnGlyLysAspGlnHisPheProValPhe 847



Mon Nov 25 08:18:23 2002

## JOURNAL

REFERENCE  
AUTHORS  
TITLE

## COMMENT

Db 2571 ACAATACCAAGTCGAATCTCATCAGACAGGCAAAACACCTTTTCCTGTAGTC 2630  
 Qy 848 MetAsnGluYgluAepIleLeuTyrCysThrGluMetGluArgValPheGlyPhePro 867  
 Db 2631 ATGAATGGCAAGACGACGTTTGTGTGCTGAGCTCGAAAGGACTTCGGCTTCCT 2690  
 Qy 868 ValHisTyrThrAspValSerAsnMetSerArgLeuAlaArgGlnArgLeuGlyArg 887  
 Db 2691 GCTCACTACAGACGCTGTCAACATGGCGCGCGCCGCTCAGAAGCTGTGGCAGG 2750  
 Qy 889 SerTrpSerValProValIleArgHisLeuPheAlaProLeuYgluTyrPheAlaCys 907  
 Db 2751 TCTGTGAGTGATACCGGTGCATCAGACACCTGTTTCCCGCTTGAAGGACTTTCCTGT 2810  
 RESULT 13  
 AC112586/c  
 LOCUS  
 DEFINITION  
 AC112586 176697 bp DNA linear HTG 13-JUL-2002  
 Rattus norvegicus clone CH230-28122, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\*, 65 unordered pieces.  
 AC112586 3 GI:21738467  
 HTG; HTGS\_PHASE1.  
 Norway rat.  
 SOURCE  
 Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 176697)  
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,  
 Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Ayele,M., Banks,T.,  
 Barbarta,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,  
 Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,  
 Buhaq,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
 Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Gorrill,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,  
 Homai,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.B.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissege,H.,  
 Lozada,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,B.,  
 Massey,E., Mahoney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,  
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,  
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,  
 Oragunye,N., Owiedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,  
 Rives,M., Rojas,A., Rojubenkan,I., Rolfe,M., Ruiz,S., Savary,G.,  
 Scherzer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I.,  
 Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H., Tang,H.,  
 Tansey,J., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Thomas,S.,  
 Toney,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Umani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,  
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,  
 Williams,G., Williamson,A., Wleczek,R., Wooden,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstein,G., and Gibbs,R.  
 Direct Submission  
 Unpublished  
 Reference 2 (bases 1 to 176697)  
 Worley,K.C.  
 Direct Submission

Submitted (22-FEB-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 176697)  
 Worley,K.C.  
 Direct Submission  
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 12, 2002 this sequence version replaced gi:20303171.

Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GRAT  
 Center clone name: CH230-28122  
 ----- Summary Statistics  
 Sequencing vector: Plasmid  
 Chemistry: Dye-terminator Big Dye: 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 121959 bases at least Q40  
 Consensus quality: 126371 bases at least Q30  
 Consensus quality: 130428 bases at least Q20

\* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 65 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1415: contig of 1415 bp in length  
 1416 1515: gap of unknown length  
 1516 2865: contig of 1450 bp in length  
 2866 3065: gap of unknown length  
 3066 4228: contig of 1163 bp in length  
 4229 4328: gap of unknown length  
 4329 5385: contig of 1057 bp in length  
 5386 5486: gap of unknown length  
 5486 6670: contig of 1185 bp in length  
 6671 6771: gap of unknown length  
 6771 7813: contig of 1043 bp in length  
 7814 7914: gap of unknown length  
 7914 9674: gap of unknown length  
 9674 11471: contig of 1697 bp in length  
 11471 11570: gap of unknown length  
 11571 13162: contig of 1592 bp in length  
 13163 14504: contig of unknown length  
 14505 14604: gap of unknown length  
 14605 15974: contig of 1370 bp in length  
 15975 16075: gap of unknown length  
 16075 17397: contig of 1322 bp in length  
 17397 19323: contig of 1827 bp in length  
 19324 21041: contig of 1618 bp in length  
 21042 21141: gap of unknown length  
 21142 22419: contig of 1278 bp in length  
 22420 22520: gap of unknown length  
 22520 23942: contig of 1423 bp in length  
 23943 24042: gap of unknown length  
 24043 25334: contig of 1291 bp in length  
 25334 25434: gap of unknown length  
 25434 26531: contig of 1097 bp in length  
 26531 26631: gap of unknown length  
 26631 28270: contig of 1640 bp in length

TITLE  
 JOURNAL  
 Unpublished  
 Reference 2 (bases 1 to 176697)  
 Worley,K.C.  
 Direct Submission

\* 28271 28370: gap of unknown length  
\* 28371 28970: contig of 1600 bp in length  
\* 29971 30070: gap of unknown length  
\* 30071 32265: contig of 2195 bp in length  
\* 32266 32366: gap of unknown length  
\* 32366 34179: contig of 1814 bp in length  
\* 34180 34279: gap of unknown length  
\* 34280 36392: contig of 2113 bp in length  
\* 36393 36493: gap of unknown length  
\* 36493 38552: contig of 2060 bp in length  
\* 38552 40932: gap of unknown length  
\* 40933 41032: contig of 2280 bp in length  
\* 41033 43021: contig of 1989 bp in length  
\* 43022 43121: gap of unknown length  
\* 43122 44977: contig of 1856 bp in length  
\* 44978 47358: gap of unknown length  
\* 47359 47554: gap of unknown length  
\* 47554 49554: contig of 2096 bp in length  
\* 49555 49654: gap of unknown length  
\* 49655 51876: contig of 2222 bp in length  
\* 51877 51976: gap of unknown length  
\* 51977 53794: contig of 1818 bp in length  
\* 53795 53894: gap of unknown length  
\* 53895 55634: contig of 1740 bp in length  
\* 55635 55734: gap of unknown length  
\* 55735 57445: contig of 1711 bp in length  
\* 57446 57545: gap of unknown length  
\* 57546 60210: contig of 2665 bp in length  
\* 60211 60310: gap of unknown length  
\* 60311 62174: contig of 1864 bp in length  
\* 62175 62274: gap of unknown length  
\* 62275 64285: contig of 2011 bp in length  
\* 64286 66764: gap of unknown length  
\* 66765 66864: gap of unknown length  
\* 66865 69398: contig of 2534 bp in length  
\* 69399 71451: gap of unknown length  
\* 71452 71551: gap of unknown length  
\* 71552 74405: contig of 2854 bp in length  
\* 74406 74505: gap of unknown length  
\* 74506 76586: contig of 2081 bp in length  
\* 76587 80397: contig of 3711 bp in length  
\* 80398 80498: gap of unknown length  
\* 80499 82399: contig of 1802 bp in length  
\* 82300 82399: gap of unknown length  
\* 82400 84796: contig of 2397 bp in length  
\* 84797 84896: gap of unknown length  
\* 84897 86981: contig of 2085 bp in length  
\* 86982 87081: gap of unknown length  
\* 87082 89794: contig of 2713 bp in length  
\* 89795 89894: gap of unknown length  
\* 89895 91696: contig of 1802 bp in length  
\* 91697 91796: gap of unknown length  
\* 91797 94458: contig of 2662 bp in length  
\* 94459 94558: gap of unknown length  
\* 94560 98352: contig of 3794 bp in length  
\* 98353 98452: gap of unknown length  
\* 98453 101134: contig of 2682 bp in length  
\* 101135 101234: gap of unknown length  
\* 101235 104588: contig of 3354 bp in length  
\* 104589 104688: gap of unknown length  
\* 104689 107761: contig of 3073 bp in length  
\* 107762 107861: gap of unknown length  
\* 107862 111340: contig of 3479 bp in length  
\* 111341 111440: gap of unknown length  
\* 111441 115774: contig of 3934 bp in length

Alignment Scores: 2.04e-88 Length: 176697  
Pred. No.:

Score: 2200.50 Matches: 415  
Percent Similarity: 89.86% Conservative: 19  
Best Local Similarity: 85.92% Mismatches: 39  
Query Match: 44.83% Indels: 11  
DB: 2 Gaps: 2  
US-09-720-086-5 (1-908) x AC112586 (1-176697)

QY 391 SerGlyLysAlaValGluValGlnAsnLysGlnMetIleGluThrPheLeuGlyGlyPhe 410  
Db 23941 TCTCGCAACGGTGTGGAGGTGATAACTA-TTGGCAATTGAGGGGTCTTGGCGGCTTC 23883  
QY 411 GlnProSerGlyProLysGlyLeuGluProProGluGluGluLysAsnProTyrIleGlyGlu 430  
Db 23882 CAGCCCTCTGTGTCACAGGCGCTGTAGCCACCAATAAGAGAGAACATCTTACAGGAA 23823  
QY 431 ValTyrThrAspMetTyrValGluProGluAlaAlaTyrAlaProProProAla 450  
Db 23822 GTTTACACCGACATGTGGGTGAGGCTGAGGCGAGCTGTATTCCCGCCACACAGCC 23763  
QY 451 LysLysProArgLysSerThrThrGluLys-ProLysValLysGluIleIleAspGluLys 470  
Db 23762 AAGAAACCCAGAAAGAGC-----ACAACTTAAGGTCAAGAGATCATTTGATGAAACA 23712  
QY 470 GThrArgGluArgLeuValTyrGluValArgGluLysCysArgAsnIleGluAspIleCy 490  
Db 23711 CCCAAGATTACGACTCATGTATGAGGTGGCCCAAAAGTCCAAAGCATCGAGACATTGG 23652  
QY 490 GlleSerCysGlySerLeuAsnValThrLeuGluHisProLeuPheIleGlyGlyMerCy 510  
Db 23651 TATCTCATGTGGGAGGCTTCATAGTCAACCTGAGACACCACCTCTTCAATTGGTGGATGTG 23592  
QY 510 GcLnaCysLysAsnCysPheLeuGluCysAlaTyrGlnTyrAspAspAspGlyTyrG1 530  
Db 23591 CCACAACTGTAAAGAACTGCTTTTGAAGTGTGCTTACCAATACATGACATGTGTACCA 23532  
QY 530 nSerTyrCysThrIleCysCysGlyGlyArgGluValLeuMeCysGlyAsnAsnAsnCys 550  
Db 23531 GTCTTCTGTACCACTCTGCTGTGGGGCGGAGTGTCTCATGTGTGGAAACAACACCGC 23472  
QY 550 eCysArgCysPheCysValGluCysValAspLeuLeuValGlyProGlyAlaAlaGlnAl 570  
Db 23471 CTGCAGGTGCTTTGTGTGTGGAGTGTGTGATCTTGTGTGGGCGCAGGGGCTGCCCAAGC 23412  
QY 570 AlaIleLeuGluAspProTyrAsnCysTyrMetCysGlyHisIleSerGlyThrTyrGlyLe 590  
Db 23411 AGCCATTAAAGAAAGACCCCTGTAACTGCTACATGTGT-----GAGCT 23370  
QY 590 uLeuArgArgArgGluAspTyrProSerArgLeuGlnMetPhePheAlaAsnAsnHisAs 610  
Db 23369 GCTGCCGAGAGAGGAGAGCTGGCTTCTCGACTCCACATGTTCTTGGCCAAATACACCA 23310  
QY 610 pGlnGluPheAspProProLysValTyrProProValProAlaGluLysArgLysProI1 630  
Db 23309 CCAGAAATTGACCCCGGAGAGTTTACCCACCTGTGTCAGCTGAGAGAGAGAAACCCAT 23250  
QY 630 eArgValIleuSerLeuPheAspGlyIleAlaThrGlyLeuLeuValLeuLysAspLeuG1 650  
Db 23249 CCACGCTACTACTATTGATGAGGATGCTACAGGCGTCCGTGCTGTAAGAGACTGGG 23190  
QY 650 yIleGlnValAspArgTyrIleAlaSerGluValCysGluAspSerIleThrValGlyWe 670  
Db 23189 CATTCAGTGGACCGCTACATGCTTCCGAGGTGCGAGGACTCCATCACCGTGGCGAT 23130  
QY 670 tValArgHisGlnGlyLysIleMetTyrValGlyAspValArgSerValThrGlnLysHi 690  
Db 23129 AGTGGCGGACACAGGAGAAAGATCATGTGTGCGGGGACGTCTGACAGATCACACAGAA 23070  
QY 690 sIleGlnGluTyrGlyProPheAspLeuValIleGlyGlySerProCysAsnAspLeuSe 710  
Db 23069 TATCCAGAGAGGGGCGCCATACATCTGTGATTTGGGGGACATCTTGCATGACCTCTC 23010  
QY 710 rIleValAsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPhePheGluPh 730

```
Db 23009 CATCATCAACCCCTGCCGCAAGGACTTTACAAGGGCACTGGCAGCCCTCTTCTTGAGTT 22950
Qy 730 eTVrAgluLeuHieAspAlaArgProLysGluGlyAspArgProPhePheTrpLe 750
Db 22949 CTACCACTCTCATGATGCGCGGCCCAAGGAGGAGGATCGCCCTTCTTCTGGCT 22890
Qy 750 uPheGluAnValValAlaMetGlyValSerAspLysArgAspLysSerArgPheLeuGl 770
Db 22889 CTTTGAAGAGGTGGTGGCCATGGCCGTTAGTGAAGAAGGAGCATCTCGGATTTCTTGA 22830
Qy 770 uSerAnProValMetIleAspAlaLysGluValSerAlaAlaHisArgAlaArgTyRPh 790
Db 22829 GTCCTAAGCTGTGATGATGATCTCAAGAAAGTGTCTGCTGCACACAGGTCGCTTACTT 22770
Qy 790 eTrpGlyAnLeuProGlyMetAnArgProLeuAlaSerThrValAnAspLysLeuGl 810
Db 22769 CTGGGGTAACCTTCTTGGCATGAACAGGCCATTGGCATCCACTATGAATGATAAGCTGA 22710
Qy 810 uLeuGlnGluCysLeuGluHisGlyArgIleAlaLysPheSerLysValArgThrIleTh 830
Db 22709 GCTGCAAGAGTGTCTGGAACATGCGACAGACAGCCAGGTTACGAAAGTGAGGACCATTAC 22650
Qy 830 rThrArgSerAnSerIleLysGlnGlyLysAspGlnHisPheProValPheMetAnGl 850
Db 22649 CACCAGGTCAAACTCGTAAGAGGAGGCAAGACAGCATTTCCCTGCTTCTCATGAATGA 22590
Qy 850 uLysGluAspIleLeuTrpCysThrGluMetGluArgValPheGlyPheProValHisTy 870
Db 22589 GAAGGAAGACATCTCTGTGTCATGAAATGGAACAGTGTGTGCTTACCTGCTCTAC 22530
Qy 870 rThrAsp 872
Db 22529 TACAGAA 22523

RESULT 14
AF331857
LOCUS Homo sapiens DNA cytosine methyltransferase 3 beta (DNMT3B) mRNA,
DEFINITION complete cds.
ACCESSION AF331857
VERSION AF331857
KEYWORDS AF331857.1 GI:18033254
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4335)
AUTHORS Ni, J., Pradhan, S. and Roberts, R. J.
TITLE Cloning, expression and characterization of human DNMT3 genes
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4335)
AUTHORS Ni, J., Pradhan, S. and Roberts, R. J.
TITLE Direct Submission
JOURNAL Submitted (22-DEC-2000) New England Biolabs, 32 Tozer Road,
Beverly, MA 01915, USA
FEATURES
source
1. .4335
/organism="Homo sapiens"
/db_xref="taxon:9606"
gene 1_4335
CDS 322..2883
/gene="DNMT3B"
/codon_start=1
/product="DNA cytosine methyltransferase 3 beta"
/protein_id="AAL57040.1"
/db_xref="GI:18033255"
/translation="MKGDTRLHNGEEDAGGREDISLVNGACSDQSDSPPTLEAIRTP
ETRGRSSRLSKREVSSLLSYTDLTGDDGDDGSDTPVPMKLFRETRRSPPA
VTRNNNSVSRERHPSRSTGRGROGNNHVDSPVFPATRSLSLRRTASACTPWS
PSSYLITLIDTDETHGTPOSSSTPYARLAODSQGGMESPOVEADSGDSEYQ
DGKEFGDLVWGKIKGFSWFPANVSVSWKATSKRQAMSGMRWQVFGDGKFSVSDAK
```

BASE COUNT 1045 a 1172 c 1167 g 951 t

ORIGIN

Alignment Scores:

Pred. No.:	5,03e-90	Length:	4335
Score:	2200.00	Matches:	461
Percent Similarity:	60.58%	Conservative:	123
Best Local Similarity:	47.82%	Mismatches:	265
Query Match:	44.82%	Indels:	117
DB:	9	Gaps:	19

US-09-720-086-5 (1-908) x AF331857 (1-4335)

```
Qy 30 ArgGlyLysGluGluArgGlnGluProSerAlaThr---AlaArgLysValGlyArgPro 48
Db 76 CGTGGAGCCTCCGAGCGCCGCCCGCAGCAGCGGACCGGCTCCTCGCGTGGGCGGAGCG 135
Qy 49 GlyArgLysArgLysHisProProValGluSerSerAspThrProLysAspProAlaVal 68
Db 136 GCGCGCAACGTCGCCGCGCGCGCTGGGGTTAAGTGGCCCAAGTAACCTAGCTCG 195
Qy 69 -----ThrThrLysSerGlnProMetAlaGluAspSerGlyProSer 82
Db 196 GCGATCGGCGCGGAGATTCGCGAGCCCGCTTCACGCGCCGCCCGCGCGCTCCCG 255
Qy 83 AspLeuLeuProAnGlyAspLeuGluLysArgSerGluProGlnProGlu----- 99
Db 256 CCAGCCAGCCCGCCCGCGCTCGCGCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCAG 315
Qy 100 -----GluGlySerProAlaAlaGlyGlnLys--- 108
Db 316 GAAAGCATGAAGGAGACACCCAGGCATCTCAATGGAGAGGAGGACCGCGCGGAGGAA 375
Qy 109 -----GlyGlyAlaProAlaGluGlyGluGlyThrGluThrProPro--- 122
Db 376 GATCGATCTCTGTCACCGGGGCTTCGCGAC---CAGTCTCCGACTCGCCCAATC 432
Qy 123 ---GluAla-----Ser 125
Db 433 CTGAGGCTATCCGCAACCCCGGAGATCAGAGCGCAGATCAAGCTCGCGACTCTCCAAG 492
Qy 126 ArgAlaValGluAnGlyCysValThrLysGluGlyArgGlyAlaSerAlaGlyGlu 145
Db 493 AGGAGGTGTCAGTCTGCTTAAGCTACACAGGACTTGACAGGCGATGGCGAGGGAA 552
Qy 146 GlyLysGluGlnLysGlnThrAsnIle-----GluSerMetLys 158
Db 553 GATGGGATGGCTCTGACACCCCGCTATGCCAAAGCTCTTCGCGGAAACAGGACTCGT 612
Qy 159 MetGluGlySerArgGlyArgLeuArgGlyGlyLeuGlyTrpGluSerSerLeuArgGln 178
Db 613 TCAGAAAGCCCGCTGTCGCACTCGAAATAACAACACAGTGTCTCCAGCCGCGAGAGCAC 672
Qy 179 ArgProMetProArgLeuThrPheGlnAlaGlyAspProTyTyRileSerLysArgLys 198
Db 673 AGGCTTCCCGACGTTCCACCC--GAGCGCGGC-----AGGGCC 709
Qy 199 ArgAspGluTrpLeuAla---ArgTrp----- 206
Db 710 GCAACCATGTGGAGCAGCTCCCGCTGGAGTTCCTGAGTCCCTGAGACGGCGGG 769
Qy 207 ---LysArgGluAlaGluLysAlaLysValIleAlaValMetAsnAlaValGluGlu 225
```

Db 770 CAACAGCATGGGACAGAACCCATGGCGCTCCCTCCAGCTCTTACTTACCATGACC 829  
QY 226 AsnGlnAlaSerGlyGluSerGlnLysValGluGlnAlaSerProProAlaValGlnGln 245  
Db 830 TCACAGACG-----ACACAGAGGACACACATCGGACGCCCCAGAGCAGCATACC 880  
QY 246 ProThrAspProAlaSerProThrValAlaThr-----ThrProGluProVal 261  
Db 881 CCAACG---CCCGCTAGCCAGGACAGCCAGAGGGGGGCGATGAGTCCCGCGAC-GTG 936  
QY 262 GlyGlyAspAlaGlyAspLysAsnAlaThrLysAlaAlaAspAspGluProGluTyrGlu 281  
Db 937 GAGCGACACAGTCGAGAT-----CGAGCACGTCACAGATTCAG 975  
QY 282 AspGlyArgGlyPheGlyIleGlyGluLeuValTyrGlyLysLeuArgGlyPheSerTyr 301  
Db 976 GATGGGAGAGAGTTTGGAAATAGGGGACCTCGTGGGAGAAAGATCAAGGCTTCTCTGG 1035  
QY 302 TyrProGlyArgIleValSerTyrTyrMetThrGlyArgSerArgAlaAlaGluGlyThr 321  
Db 1036 TGGCCCCCATAGTGGTGTCTTGGAAAGCCACCTCCAGCGACAGGCTATGTCTGCGATG 1095  
QY 322 ArgTyrValMetTyrPheGlyAspGlyLysPheSerValValCysValGluLysLeuMet 341  
Db 1096 CGGTGGGTCCAGTGGTGTGGCGATGCAAGTCTCCGAGGTCTCTGACAGCAAACTGATG 1155  
QY 342 ProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnProMetTyr 361  
Db 1156 GCACTGGGGCTGTTCAGCCAGCACCTTAAATTGGCCACCTTCAATTAAGCTCTCTCTAT 1215  
QY 362 ArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeuPhePro 381  
Db 1216 CGAAACCATAGTACATGCTCTGGAAGAGCTAGGGTGGAGCTGGCAAGACCTTCCCC 1275  
QY 382 AlaCysHisAspSerAspGluSerAspSerGlyLysAlaValGluValGlnAsnLysGln 401  
Db 1276 -----AGCAGCCCTGGAGACTCATTTGAGAGGACCGAGCTGAAGCCC 1314  
QY 402 MetIleGluTyrAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGluProPro 421  
Db 1315 ATGTGGAGTGGGCCACAGGGGGCTTCAAGCCACTGGATCGAGGGCTCAAAACCAAC 1374  
QY 422 GluGlu-----Glu 424  
Db 1375 AACACGCAACGAGTGAATTAAGTCGAAGGTGCTGTGCGAGCGAGTAGAAATTAAGAA 1434  
QY 425 LysAsnProTyrLysGluValTyrThrAspMetTyrValGluProGlnAlaAla--- 443  
Db 1435 TCAAGAAATAGCAGAAACAAGACTCGAAGACGACAGCTGCGACTCCACTCCGTGAC 1494  
QY 444 TyrAlaProProProProAlaLysLysProAlaGlySerThrThrGluLysProLysVal 463  
Db 1495 TACTGCCCGCGACCCCAAGCCCTCAAGCAAAATTGCTATAACAACGCGCAAGACGAGG 1554  
QY 464 LysGluIleIleAspGluArgThrArgGluArgLeuValTyrGluValAlaArgLysCys 483  
Db 1555 GATGAA-----GATCAGAGCGGAGAAACAATGGCTTCAGATGTTGCCAACAAACAG 1605  
QY 484 ArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThrLeuGluHisPro 503  
Db 1606 AGCAGCCCTGGAAGATGGCTGTGTCTGTGCGAGGAAAAACCCCGTCTCTTCCACCT 1665  
QY 504 LeuPheIleGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGluCysAlaTyrGln 523  
Db 1666 CTCTTTAGAGGGGGGCTCTGTCAAGCATGCGGGATGCTTCTTGAGGTGTTTAAATG 1725  
QY 524 TyrAspAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyGlyArgGluValLeu 543  
Db 1726 TATATGATGCGATGGCTATAGTCTTAATGCACTGTGTGCTGCGAGGCGGAGCTGCTG 1785  
QY 544 MetCysGlyAsnAsnAsnCysCysArgCysPheCysValGluCysValAlaAspLeuVal 563  
Db 1786 CTTTGACGCAACAGAGCTGTGCTGGGTGTTTCTGTGTGAGGTGCTGGAGGTGCTGCTG 1845

QY 564 GlyProGlyAlaAlaGlnAlaIleLysGluAspProTyrAsnCysTyrMetCysGly 583  
Db 1846 GGCACAGGCGACAGCGGCGGAGCCCAAGCTTCCAGAGAGCCCTGGAGCTGTTCATGTCTTC 1905  
QY 584 HisLysGlyThrTyrGlyLeuLeuArgArgGluAspTyrProSerArgLeuGlnMet 603  
Db 1906 CCGCACGCTGTCAATGAGCTCTGCGCGCGGAGGAGCATGGAACGCGCGCTCGAGGCC 1965  
QY 604 PhePheAlaAsnAsnHisAspGlnGluPheAspProProLysValTyrProProValPro 623  
Db 1966 TTCCTTCAACGATGACAGCGGGCTGTGAATATGAAGCCCCCAAGCTGATCCCTGCC 2025  
QY 624 AlaGluLysArgLysProIleArgValLeuSerLeuPheAspGlyIleAlaThrGlyLeu 643  
Db 2026 GCAGCCCGAAGCGCGCCATTTGAGTCTGATTTGATGGCATCGCAGACGCTAC 2085  
QY 644 LeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSerGluValCysGlu 663  
Db 2086 CTAGCTCTCAAGAGTGGGCAATTAAGGTAGGAAGTACGTGCTCTGAAAGTGTGAG 2145  
QY 664 AspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyrValGlyAspVal 683  
Db 2146 GAGTCCATTCGTGTGGAAACCGGAAGCAGGAGGAAATATCAAAATCAAGCAACGCTG 2205  
QY 684 ArgSerValThrGlnLysHisIleGlnGluTyrProPheAspLeuValIleGlyGly 703  
Db 2206 AGGAACATCAACAAAGAAATATTAAGAAATGGGGCCCATTTGACTGGTATTTGGCGA 2265  
QY 704 SerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeuTyrGluGlyThr 723  
Db 2266 AGCCCAAGCAAGCATCTCTCAATGTGAATTCAGCCAGGAAAGCCGTATGAGGGTACA 2325  
QY 724 GlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgProLysGlyAsp 743  
Db 2326 GGCGGCTCTCTTCTGAAATTTTACACCTGCTGAATTTACTACGCCCCCAAGAGGGTAT 2385  
QY 744 AspArgProPhePheTyrLeuPheGluAsnValAlaMetGlyValSerAspLysArg 763  
Db 2386 GACCGGCGCTTCTGTGATGTTGAGAAATGTTGACCATGAAGTTGGCGACAGAGG 2445  
QY 764 AspIleSerArgPheLeuGluLeuAsnProValMetIleAspAlaLysGluValSerAla 783  
Db 2446 GACATCTCACGGTTCCTGAGAGTGAATCCAGTGAATGATGATCCATCAAGTTCTGCT 2505  
QY 784 AlaHisArgAlaArgTyrPheTyrPheTyrAsnLeuProGlyMetAsnArgProLeuAlaSer 803  
Db 2506 GCTCACAGGGCCGATCTTCTGGGCAACCTTACCCGGGATGAACAGGCCGTGATAGCA 2565  
QY 804 ThrValAsnAspLysLeuGluLeuGlnGluCysLeuGluHisGlyArgIleAlaLysPhe 823  
Db 2566 TCAAGAAATGATTAACCTGAGCTGACAGACTGTGGAATTAACAATAGCATACCAAGTTA 2625  
QY 824 SerLysValArgThrIleThrThrArgSerAsnSerIleLysGlnGlyLysAspGlnHis 843  
Db 2626 AAGAAAGTACAGACATTAACACCAAGCTGAATCGATCAACAGGGGAAACCAACTT 2685  
QY 844 PheProValPheMetAsnGluLysGluAspIleLeuTyrCysThrGluMetGluArgVal 863  
Db 2686 TTCCTGTTCATGATATGCAAGAGATGTTTGTGTGCTGACTGAGCTCAAAAGATC 2745  
QY 864 PheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeuAlaArgGlnArg 883  
Db 2746 TTGGCTTTCTGTGACATCAACAGAGCTGTTCACACATGGCGCGTGGCGCCGAGAG 2805  
QY 884 LeuLeuGlyArgSerTyrPheValProValIleArgHisLeuPheAlaProLeuLysGlu 903  
Db 2806 CTGCTGGGAAGTCTGTGAGCGGTGCTGTCTCATCCAGACACTTTCGCCCTCTGAAGGAC 2865  
QY 904 TyrPheAlaCys 907  
Db 2866 TACTTTGCATGT 2877

RESULT 15

AF151973 4223 bp mRNA linear ROD 08-JUN-2000  
 LOCUS Mus musculus DNA cytosine-specific methyltransferase isoform 5  
 DEFINITION (Dnmt3b) mRNA, complete cds.  
 ACCESSION AF151973  
 VERSION AF151973.1 GI:8347127  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 4223)  
 YIN, B., CHEN, Y. T., ZHU, M., LUO, Y. J., ZHU, N., XU, S. C., WU, G. Y. and  
 SHEN, Y.  
 TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing  
 isoforms in mouse embryonic tissue  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 4223)  
 YIN, B., CHEN, Y. T., ZHU, M., LUO, Y. J., ZHU, N., XU, S. C., WU, G. Y. and  
 SHEN, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular  
 Biology, Institution of Basic Medical Sciences, Chinese Academy of  
 Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,  
 P.R.China

## FEATURES

source Location/Qualifiers  
 1..4223  
 /organism="Mus musculus"  
 /strain="KM"  
 /db\_xref="taxon:10090"  
 /dev\_stage="8-9 day old embryo"  
 1..4223  
 /gene="Dnmt3b"  
 297..2876  
 /gene="Dnmt3b"  
 /function="methylates cytosine in DNA"  
 /note="alternatively spliced"  
 /codon\_start=1  
 /product="DNA cytosine-specific methyltransferase isoform 5"  
 /protein\_id="AAF74519.1"  
 /translation="MKGPSRLHNEBEGASGVEEIIVNGNFSQSDTKDAPSPVLE  
 AICTPEVCTPRTGRSSRLSKREVSSLLVNTQMDTGDDRDDEVDVNGSILMPK  
 LTRTKQTRTRESAPVATRHNSGTSLSERQASPRITRGRGRHVRHVEYTPVFPATR  
 SRRRRASSTPWSSPASVDPEEVTIPKSVSTPSVDSQDQEGMDTTPQVDAESIY  
 GDSIEYQDDKEFGIDLVWVKIKGFSWPMVVMVSMKATSKRQMPGMWQVDFGDKF  
 SEISADLVGLFQHFNLATFNKLVSRYKAMHYHLEKARVAGRTFFSSPGESLED  
 QLKPMLEWAHGGFKPTGIEGLKPNKQPVNKSVRSDSRNLEPRRNSRRRTN  
 DSAASESPPPKRLKNTSYGKDRGEDESRERMASEVTNNKGLNDRCLSCGKNPV  
 FHLFEGGLCCSDRDLFLFYMDGYSYTCCEGRELLCNTSCRCFCVCE  
 LEVLVAGTAEADAKLOEPWCVMCLPQCHGLRRKDKNMRLODFETDPDLEREP  
 PKLYPAIAKRRPIRLVSLPDGIATGLVLEKLGKVEKYIASEVCAESIAVGTVKH  
 EGOIKYVNDVRKIYKKNIEEWGPFDLVIGGSPCNLSNVNPAKKGLEYGTGRUFFEY  
 HLNLYTRKEGDNRPFFMWFENVAMKVNDDKDIKFLACNPNPMIDAIKVSAARHRY  
 FWNGLPMGNRPVMASKNDKLEQDFSRFATLKKVOTITTKNSIRQGNQLFPVY  
 MNGKDDVLWCTELERIFGPPAHYTDVNMGRARQKLGRSWSVPVIRHLFAPLKDYF  
 ACE"

BASE COUNT 1051 a 1084 c 1147 g 941 t

## ORIGIN

Alignment Scores:  
 Pred. No.: 5, 72e-90 Length: 4223  
 Score: 2198.50 Matches: 449  
 Percent Similarity: 60.42% Conservative: 131  
 Best Local Similarity: 46.77% Mismatches: 238  
 Query Match: 44.79% Indels: 143  
 DB: 10 Gaps: 15

US-09-720-086-5 (1-908) x AF151973 (1-4223)

QY 44 LysValGlyArgProGlyArgLysArgLys-----53

Db 130 CGACGGGCGCTCCAGCTCCGGCAGCGCGGGTTAAGCGGGCCCAAGTAAACGTAGCGCAG 189  
 QY 54 -----HisProProValGluSerSerAspThrProLysAspProAlaValThrThrLys 71  
 Db 190 CGATCGCGCGCGGAGATTCCGGAACCCGACACTCGCGCGCGCGCGCGCGCGCGCGCG 249  
 QY 72 -----SerGlnProMetAlaGlnAspSerGlyProSerAsp 83  
 Db 250 GCGCGGATCCGCGCGCGCGCTACAGCCAGCTCA-CGACAGGAACAATGAAGGAGAC 308  
 QY 84 -----LeuLeuProAsn 87  
 Db 309 AGCAGACATCTGAATGAAGAAGAGGGTCCAGCGGTATGAGGAGTGCATTATCGTTAAT 368  
 QY 88 GlyAspLeuLysArgSerGluProGlnProGluGluGlySerProAlaAlaGlyGln 107  
 Db 369 GGAACCTTCAGTACCAGCTCTCAGACACGAAAGATGCTCCCTCACCC-----416  
 QY 108 LysGlyGlyAlaProAlaGluGlyGlyThrGluThrProGluAlaSerArgAla 127  
 Db 417 -----CCAGTCTTGAGGCAATCTGC 437  
 QY 128 ValGluAsnGlyCysCysValThrLysGluGlyArgGlyAlaSerAla-----143  
 Db 438 ACAGACCCAGTCTGCACACAGACAGCCAGGCGCGAGTCAAGCTCCCGGTGTCTAAG 497  
 QY 144 -----GlyGluGly-----146  
 Db 498 AGGGAGGTCTCCAGCTTCTGAATTACACGACGACATGACAGGAGATGGACACAGAT 557  
 QY 147 -----LysGluGln 149  
 Db 558 GATGAAGTAGATGATGGGAATGGCTGTGATATTCTAATGCCAAGCTCACCGTGAGACC 617  
 QY 150 LysGlnThrAsnIleGluSerMetLysMetGluGlySerArgGlyArgLysArgGlyGly 169  
 Db 618 AAGGACACCGAGCGGCTCT-----GAAAGCGCGGTGTCCGAACCCGACATAGC 668  
 QY 170 LeuGlyTrpGluSerSerLeuArgGlnArgProMetProArgLeuThrPheGlnAlaGly 189  
 Db 669 AATGGGACCTCCAGCTTGGAGAGGCAAGAGCCCTCCCGCAGAAATCACCGAGTCCGGCAG 728  
 QY 190 AspProTyTrpLysSerLysArgLysArgAspGluTrpLeuAlaArgTrpLysArgGlu 209  
 Db 729 GCGCGCACCATGTG-----CAGGATACCTGTGGAGTTTCCGGCTACC 773  
 QY 210 AlaGluLysLysAlaLysValIleAlaValMetAsnAlaValGluGluAsnGlnAlaSer 229  
 Db 774 AGGTCTCGGAGCGTCGAGCATCGTCTTTCAGCAGCAGCCATGGTCATCCCTCCCGCAGC 833  
 QY 230 GlyGluSerGlnLysValGluGluAlaSerProProAlaValGlnGlnProThrAspPro 249  
 Db 834 GTCGACTTC-----ATGGAAGAAAGTGACACCTAAGAGCGTCAGTACCCCA-----878  
 QY 250 AlaSerProThrValAlaThrThrProGluProValGlyGlyAspAlaGlyAspLysAsn 269  
 Db 879 ---TCAGTTGACTTGAGCCAGGATGGAGATCAGGAGGTTATGATACACACAGGTGGAT 935  
 QY 270 AlaThrLysAlaAlaAspAspGluProGluTyTrpGluAspGlyArgGlyPheGlyIleGly 289  
 Db 936 GCAGAGACATATATGGAGACAGCAGAGATATCAGATGATAAAGAGTTGGAAATAGGT 995  
 QY 290 GluLeuValTrpGlyLysLeuArgGlyPheSerTrpTrpProGlyArgLysValSerTrp 309  
 Db 996 GACCTCGTGTGGGAAAGATCAAGGGCTTCTCTGTGGTGGCTGCCATGGTGTGTCTCTGG 1055  
 QY 310 TrpMetThrGlyArgSerArgAlaAlaGluGlyThrArgTrpValMetTrpPheGlyAsp 329  
 Db 1056 AAAGCCACCTCCAGGACGAGCCATGCCCGGANTGCCGTGGGTACAGTGTGTGTGGTAT 1115  
 QY 330 GlyLysPheSerValValCysValGluLysLeuMetProLeuSerSerPheCysSerAla 349

```

Db 1116 GGCAGATTCTTGAGATCTCTGCTGACAACTGTGGCTCTGGGGCTGTCACGCAC 1175
QY 350 PheHieglialaThThrsanlysglnProMetTyrArglysalalileTygluValleu 369
Db 1176 TTATAATGGCTACCTTCATATAGCTGGTTCTTATAGAAAGGCATGTACACACTCTG 1235
QY 370 GluValAlaSerSerArgalaglyLysleuPheProAlaCysHisAspSerAspGluSer 389
Db 1236 GAGAAAGCCAGAGGTTCCAGAGCTGGCAAGACTTC-----TCCAGC 1274
QY 390 AspSerArglyLysAlaValGluValGlnAsnlysglnMetilegluTTPAlaLeuGly 409
Db 1275 AGTCTTGAGAGAGTCACTGAGGAGCACAGCTGAAGCCATGTCTGAGTGGCCACAGGTGGC 1334
QY 410 PheGlnProSerArglyProlysglyLeuGluPro----- 420
Db 1335 TTCAGCCTACTCTGGGTTCGAGGGGCTCAAAACCAAGAACCAACAGGTAAATAG 1394
QY 421 -----ProGluGluGluLysAsnProTyr 428
Db 1395 TCGAAGGTGCGTCTGACAGACAGTAAGAACTTGAACCCAGAGACGCGAACAAGT 1454
QY 429 LysGluValTyrThrsMetTTPValGluProGluAlaAlaAlaTyrAlaProProPro 448
Db 1455 CGAAGACGACACCAAT-----GACTCTGCTCTTCTGAGTCCCCCA 1499
QY 449 ProAlaLysLysProArgLysSerThrThrgluLysProLysValLysGluIleIleAsp 468
Db 1500 CCCAAGGCGCTCAGCAAAATAGCTATGCGCGAAGACGAGAGGAG-----GAT 1550
QY 469 GluArgThraArgGluArgLeuValTyrGluValArgGlnLysCysArgAsnIleGluAsp 488
Db 1551 GAGAGAGCCCGAAGCGAGTGGCTTCTGAGATCACCAACCAAGGCGAATCTGGAAGAC 1610
QY 489 IleCysIleSerCysglYserLeuAsnValThrLeuGlnHisProLeuPheIleGly 508
Db 1611 CGGTGTTGTCTGTGGAAAGAAAGAACCTGTCTCTTCCACCCCTTTTGAAGGGGG 1670
QY 509 MetCysglAsnSerCysLysAsnCysPheLeuGluCysAlaTyrGlnTyrAspAspArgly 528
Db 1671 CTCTGTGAGATGCGCGGATGCTCTAGAGCTTCTTACATATATATGAGAGCGGC 1730
QY 529 TyrGlnSerTyrCysThrIleCysCysglYArgGluValLysMetCysglYAsnAsn 548
Db 1731 TATCAGTCTTACTGACCGGTGTGTGTAGGCGCGTGAACCTGCTGTGCACTAACA 1790
Q 549 AsnCysCysArgCysPheCysValGluCysValAspLeuLeuValGlyProGluAlaAla 568
Db 1791 AGCTGTGACAGATGCTTCTGTGTGAGATGTCTGAGGTGTGTGGCGCACAGCTGAC 1850
QY 569 GluAlaAlaIleLysgluAspProTyrAsnCysTyrMetCysglYHisLysglYThrTyr 588
Db 1851 GAGATGCGCAAGCTGAGGAACCTCGAGAGCTGTATATGTCTCCCTGACGCTGCAT 1910
QY 589 GlyLeuLeuArgArgArgLysAspTyrProSerArgLeuGlnMetPhePheAlaAsnAsn 608
Db 1911 GGGGTCTTCGCGCAGGAAGATGGAAACATGCGCTGCAAGACTTCTTCACTACTGAT 1970
QY 609 HisAsp---GlnGluPheAspProProLysValTyrProProValProAlaGluLysArg 627
Db 1971 CTGACCTGGAAGAAATTTGAGCCACCAAGTTGTACCCAGCAATTCCTGACGCAAAAG 2030
QY 628 LysProIleArgValLeuSerLeuPheAspGlyIleAlaThrglyLeuLeuValLys 647
Db 2031 AGGCCCATTAAGTCTCTGTGTGATGAAATTCACACGGGGTACTTGTGTCAAG 2090
QY 648 AspLeuGlyIleGlnValAspArgTyrIleAlaSerGluValCysGluAspSerIleThr 667
Db 2091 GAGTTGGGATTAAAGTGAAGATGACATGCTCCGAAGTCTGTGACAGATCATGCT 2150
QY 668 ValGlyMetValArgHisGlnGlyLysIleMetTyrValGlyAspValArgSerValThr 687
Db 2151 GTGGGAACCTTAAGCATGAAGGCCAGATCAATATGTCAATGACGTCCGAAATATCAC 2210

```

```

QY 688 GlnLysHisIleGlnGluTyrGluProPheAspLeuValIleGlyglYserProCysAsn 707
Db 2211 AAGAAAAATATTAAGAGAGGGGCCCGTTGACCTTGATGTGGTGAAGCCCAATGCAAT 2270
QY 708 AspLeuSerIleValAsnProAlaArgLysGlyLeuTyrGlnGlyThrGlyArgLeuPhe 727
Db 2271 GATCTCTTAACGTCMACTCTCCCGCAAGGTTATATGAGGGCACAGGAAGGCTCTTC 2330
QY 728 PheGluPheTyrArgLeuLeuHisAspAlaArgProLysGluGluLysAspArgProPhe 747
Db 2331 TTCGAGTTTAACCACTGTGATTAATACCCGCGCCAGAGAGGCGCAACGTCATTC 2390
QY 748 PheTyrLeuPheGluAsnValAlaMetGlyValSerAspLysArgAspIleSerArg 767
Db 2391 TCTGTGATGTTGAGAAATGTTGGCCATGAAAGTGAAGTGAAGAACATCTCAAG 2450
QY 768 PheLeuGluSerAsnProValMetIleAspAlaLysGluValSerAlaAlaHisArgAla 787
Db 2451 TTCCTGGCATGTAAACCACTGATGATGATGCCATCAAGGTGTCTGCTCACAGGGCC 2510
QY 788 ArgTyrPheTyrGlyAsnLeuProGlyMetAsnArgProLeuAlaSerThrValAsnAsp 807
Db 2511 CGGTACTTCTGGGTAACTTACCTACCGGATGAACAGGCCCGTGTGCTTCAAGATGAT 2570
QY 808 LysLeuGluLeuGlnGluCysLeuGluHisglYArgIleAlaLysPheSerLysValArg 827
Db 2571 AAGCTCAGCTGACAGAGCTGCTGAGTTCACTAGAGACAGCAAGTTAAAGAAAGTGAC 2630
QY 828 ThrIleThrThrArgSerAsnSerIleLysGlnGlyLysAspGlnHisPheProValPhe 847
Db 2631 ACATTAACCAACAGTCACTCATCAGACAGGCGCAAAACCAAGCTTTTCCCTGATGC 2690
QY 848 MetAsnGluLysGluAspIleLeuTyrProCysThrGlnMetGluArgValPheGlyPhePro 867
Db 2691 ATGAATGCAAGAGACAGCTTTGTGTGTGACTGAGCTGAAAGGATCTTGGCTTCCT 2750
QY 868 ValHisTyrThrAspValSerAsnMetSerArgLeuAlaArgGlnArgLeuLeuGlyArg 887
Db 2751 GCTCACTAACAGACAGCTGTCCAAATGATGGCGCGCGCGCTCAAGAGCTGCTGGCAGG 2810
QY 888 SerTyrSerValProValIleArgHisLeuPheAlaProLeuLysglYThrPheAlaCys 907
Db 2811 TCCTGAGTGTAAACGGGTATCAAGACACTGTTCCTGCCCCCTTGAAGGACTACTTGCCTGT 2870

```

Search completed: November 22, 2002, 10:51:23  
 Job time : 5545.7 secs

THIS PAGE BLANK (USPTO)





Li E, Okano M, Xie S;  
WPI; 2000-106298/09.  
P-PSDB; AAY54055.  
New mouse and human polypeptides, useful to treat and diagnose  
neoplastic disorders e.g. carcinomas, sarcomas and leukemias -  
Claim 8; Fig 1A; 114pp; English.  
The present sequence encodes a murine de novo DNA cytosine  
methyltransferase designated Dmmt3a. The polypeptides can be administered  
therapeutically, especially by expressing encoding polynucleotides, to  
treat diseases associated with DNA cytosine methyltransferase, such as  
neoplastic disorders e.g. carcinomas, sarcomas and leukemias. They can  
be used to diagnose, or determine susceptibility to neoplastic disorders,  
by assaying for polypeptide expression levels in mammalian cells/body  
fluids. They are useful to screen for compounds inhibiting/activating the  
polypeptide. The polypeptides can also be used for in vitro de novo  
methylation of DNA. Such in vitro methylation may be used to direct or  
regulate DNA expression in biological systems, e.g. recombinant DNA  
methylated in vitro may be introduced into a cell/organism to increase  
or decrease expression of a desired polypeptide for which the native DNA  
is under-methylated or not methylated. The polypeptides can also be  
used to produce antibodies which are useful to detect and purify the  
polypeptide or therapeutically e.g. to treat neoplastic disorders. The  
polynucleotides are useful to produce probes and primers which are  
useful diagnostically.

Sequence 4192 BP; 1095 A; 1070 C; 1182 G; 844 T; 1 other;

Query Match 99.7%; Score 4181.4; DB 21; Length 4192;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 4186; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 GAATTCGGCTGCTGCGGGCGCCGACCCGCGGCGCCACAGCGCAGCGCGCTGAA 60  
Db 1 GAATTCGGCTGCTGCGGGCGCCGACCCGCGGCGCCACAGCGCAGCGCGCTGAA 60  
Qy 61 GCCAGCGCTGAGCTGCTTTCCGAGGCTTGACATCAGGGTCTATGTTAAGTCTT 120  
Db 61 GCCAGCGCTGAGCTGCTTTCCGAGGCTTGACATCAGGGTCTATGTTAAGTCTT 120  
Qy 121 AGCTCTGTGTTACAAAGACCAAGCAATTCCTTCTGAGCCCTCGCAGCCCAAGG 180  
Db 121 AGCTCTGTGTTACAAAGACCAAGCAATTCCTTCTGAGCCCTCGCAGCCCAAGG 180  
Qy 181 CCTTCGAGCCCGCAGCTGCGGCTACTGCGCCAGCAATGCGCTCCAGCGGCCCGGGGAC 240  
Db 181 CCTTCGAGCCCGCAGCTGCGGCTACTGCGCCAGCAATGCGCTCCAGCGGCCCGGGGAC 240  
Qy 241 ACCAGAGCTCTCTGAGCGGAGGATGATCGAAGGAGGAGGAGGAGGAGGAG 300  
Db 241 ACCAGAGCTCTCTGAGCGGAGGATGATCGAAGGAGGAGGAGGAGGAGGAGGAG 300  
Qy 301 AACCGTGGCAAGGAGCGCCAGGAGCCAGCGCCAGCGCCAGCGCCAGCGCGGAGGCT 360  
Db 301 AACCGTGGCAAGGAGCGCCAGGAGCCAGCGCCAGCGCCAGCGCCAGCGCGGAGGCT 360  
Qy 361 GCGCGAAGCGGAGGAGCGCCAGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420  
Db 361 GCGCGAAGCGGAGGAGCGCCAGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420  
Qy 421 ACCACCAAGTCTCAGGCCATGGGCCAGGACTCTGGCCCTCAGATCTGCTACCCCAATGA 480  
Db 421 ACCACCAAGTCTCAGGCCATGGGCCAGGACTCTGGCCCTCAGATCTGCTACCCCAATGA 480  
Qy 481 GACTTGGAGAGCGGAGTGAACCCCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540  
Db 481 GACTTGGAGAGCGGAGTGAACCCCACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540  
Qy 541 GGTGGGGCCCGCAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600  
Db 541 GGTGGGGCCCGCAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600

Db 541 GGTGGGGCCCGCAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 600  
Qy 601 GAGAATGGCTGCTGTGTGACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660  
Db 601 GAGAATGGCTGCTGTGTGACCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 660  
Qy 661 CAGAAGCAGCAACATCGAATCCATGAAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 720  
Db 661 CAGAAGCAGCAACATCGAATCCATGAAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 720  
Qy 721 GGCTTGGGCTGGGAGTCCAGGCTCCGTCAGCGAGCCCATGTCACCAAGACTCACCTTCCAGGCA 780  
Db 721 GGCTTGGGCTGGGAGTCCAGGCTCCGTCAGCGAGCCCATGTCACCTTCCAGGCA 780  
Qy 781 GGGGACCCCTTACTACATCAGCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840  
Db 781 GGGGACCCCTTACTACATCAGCAACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840  
Qy 841 GAGGCTGAGAAGAAAGCAAGGTAAATTCAGTAATGCTGTGGAAGAGAACCAAGGCC 900  
Db 841 GATGCTGAGAAGAAAGCAAGGTAAATTCAGTAATGCTGTGGAAGAGAACCAAGGCC 900  
Qy 901 TCTGGAGAGTCTCAGAGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960  
Db 901 TCTGGAGAGTCTCAGAGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960  
Qy 961 CTTGCTTCTCCGACTGTGGGCAACCCCTGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020  
Db 961 CTTGCTTCTCCGACTGTGGGCAACCCCTGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020  
Qy 1021 AATGCTTACCAAAAGCAGCGCAGTGAAGTATGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080  
Db 1021 AATGCTTACCAAAAGCAGCGCAGTGAAGTATGAGGATGAGGAGGAGGAGGAGGAGGAGGAG 1080  
Qy 1081 GGAGAGTGTGTGGGGGAACTTTCGGGGCTTCTCTGGTGGCCAGGCGGAGGAGGAGGAGGAGGAG 1140  
Db 1081 GGAGAGTGTGTGGGGGAACTTTCGGGGCTTCTCTGGTGGCCAGGCGGAGGAGGAGGAGGAGGAG 1140  
Qy 1141 TGGTGTGATGACAGCGCCGAGCGCAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGA 1200  
Db 1141 TGGTGTGATGACAGCGCCGAGCGCAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGA 1200  
Qy 1201 GATGGCAAGTCTCAGTGGTGTGTGGAGAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGA 1260  
Db 1201 GATGGCAAGTCTCAGTGGTGTGTGGAGAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGAAGGAGTGA 1260  
Qy 1261 GCATTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320  
Db 1261 GCATTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320  
Qy 1321 CTCAGGTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380  
Db 1321 CTCAGGTGGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380  
Qy 1381 AGTGACAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440  
Db 1381 AGTGACAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440  
Qy 1441 GGCTTCCAGGCTCCGGTCTTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500  
Db 1441 GGCTTCCAGGCTCCGGTCTTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500  
Qy 1501 AAGGAAGTTTACACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560  
Db 1501 AAGGAAGTTTACACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560  
Qy 1561 CCAGCCAAAGAAACCCAGAAAGAGCAACAGAGAAACCTTAAGGTCAAGGAGATCATTTGAT 1620  
Db 1561 CCAGCCAAAGAAACCCAGAAAGAGCAACAGAGAAACCTTAAGGTCAAGGAGATCATTTGAT 1620  
Qy 1621 GAGCGGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680  
Db 1621 GAGCGGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1680

QY 1681 ATTTGATCTCATGSGAGCCTGAAATGACCCCTGGAGCACCCACTTTCATTTGAGGC 1740  
DB 1681 ATTTGATCTCATGSGAGCCTGAAATGACCCCTGGAGCACCCACTTTCATTTGAGGC 1740  
QY 1741 ATGAGCAAGAACTGTAGAACTGCTTCTTGAGTGTGCTTACCAATATGACGACATGGG 1800  
DB 1741 ATGAGCAAGAACTGTAGAACTGCTTCTTGAGTGTGCTTACCAATATGACGACATGGG 1800  
QY 1801 TACCAAGTCTATTGACCAATCTGTGTGGGGCGTGAAGTGTCTATGTGTGGAAACAC 1860  
DB 1801 TACCAAGTCTATTGACCAATCTGTGTGGGGCGTGAAGTGTCTATGTGTGGAAACAC 1860  
QY 1861 AACGCGTCAAGGTGTCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1920  
DB 1861 AACGCGTCAAGGTGTCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1920  
QY 1921 CAGGACGCAATTAAGGAAGCCCTGGAAGTGTCTATGTCGGGGCAATGAAGGCACCTAT 1980  
DB 1921 CAGGACGCAATTAAGGAAGCCCTGGAAGTGTCTATGTCGGGGCAATGAAGGCACCTAT 1980  
QY 1981 GGGCTGTGCGAAGACGGGAAGACTGTGCTTCTGCACTCCAGATGTTCTTTGCCAATAAC 2040  
DB 1981 GGGCTGTGCGAAGACGGGAAGACTGTGCTTCTGCACTCCAGATGTTCTTTGCCAATAAC 2040  
QY 2041 CATGACCAAGAAATTTGACCCCGCAAGGTTTACCCCACTGTGCCAGCTGAGAAAGAGAG 2100  
DB 2041 CATGACCAAGAAATTTGACCCCGCAAGGTTTACCCCACTGTGCCAGCTGAGAAAGAGAG 2100  
QY 2101 CCCATCGCGTGTCTGTCTCTTTGATGGATGCTACAGAGGCTCTGTGTGTGAAGGAC 2160  
DB 2101 CCCATCGCGTGTCTGTCTCTTTGATGGATGCTACAGAGGCTCTGTGTGTGAAGGAC 2160  
QY 2161 CTGGGCAATCCAAAGTGGACCGCTAATGCTCTCGAGGTGTGTGAAGGACTCCATCAGG 2220  
DB 2161 CTGGGCAATCCAAAGTGGACCGCTAATGCTCTCGAGGTGTGTGAAGGACTCCATCAGG 2220  
QY 2221 GGCATGTGTGCGGACCAAGGAAGATCATGTGTGGGACGTCGCGAGCGTCACACAG 2280  
DB 2221 GGCATGTGTGCGGACCAAGGAAGATCATGTGTGGGACGTCGCGAGCGTCACACAG 2280  
QY 2281 AAGCATATCCAGAGTGGGGCCCATTCGACTGTGATGGAAGGACGTCGCGCAATGAC 2340  
DB 2281 AAGCATATCCAGAGTGGGGCCCATTCGACTGTGATGGAAGGACGTCGCGCAATGAC 2340  
QY 2341 CTCTCCATTGTCAACCCCTGCGCAAGGACTTTATGAGGATCTGCGCCCTCTCTTT 2400  
DB 2341 CTCTCCATTGTCAACCCCTGCGCAAGGACTTTATGAGGATCTGCGCCCTCTCTTT 2400  
QY 2401 GAGTTCTACCGCTCTGATGATGCGCGGCCCAAGGAGGAGATGATCGCCCTTCTTC 2460  
DB 2401 GAGTTCTACCGCTCTGATGATGCGCGGCCCAAGGAGGAGATGATCGCCCTTCTTC 2460  
QY 2461 TGCTCTTTGAGAAATGTGTGGCCATGAGGCGTTAGTACAAAGAGGACATCTGCGAATTT 2520  
DB 2461 TGCTCTTTGAGAAATGTGTGGCCATGAGGCGTTAGTACAAAGAGGACATCTGCGAATTT 2520  
QY 2521 CTTGAGTCTAACCCCGTGTATGATGACGCAAGAGTGTCTGTGACACAGAGGCGGT 2580  
DB 2521 CTTGAGTCTAACCCCGTGTATGATGACGCAAGAGTGTCTGTGACACAGAGGCGGT 2580  
QY 2581 TACTTCTGGGGTAACTTCTCTGATGAACAGGCTTTGGCATCTGATGATGATGATG 2640  
DB 2581 TACTTCTGGGGTAACTTCTCTGATGAACAGGCTTTGGCATCTGATGATGATGATGATG 2640  
QY 2641 CTGAGAGTGAAGAGTGTCTGAGACAGGCAAGTATAGCAAGTTTCAGCAAAAGTGAAGACC 2700  
DB 2641 CTGAGAGTGAAGAGTGTCTGAGACAGGCAAGTATAGCAAGTTTCAGCAAAAGTGAAGACC 2700  
QY 2701 ATTACCAACAGTCAAACTCTATTAAGCAAGGCAAGGCAAGCAATTTCCCGCTTTCATG 2760  
DB 2701 ATTACCAACAGTCAAACTCTATTAAGCAAGGCAAGGCAAGCAATTTCCCGCTTTCATG 2760

QY 2761 AACGAGAAGAGACATCTGTGTGCACTGAATGGAAGGGGTTTGGCTTCCCGCTC 2820  
DB 2761 AACGAGAAGAGACATCTGTGTGCACTGAATGGAAGGGGTTTGGCTTCCCGCTC 2820  
QY 2821 CACTACACAGAGCTGTCCCAATGAGGCGCTTGGGAGAGACCTGTGGGCGCATCG 2880  
DB 2821 CACTACACAGAGCTGTCCCAATGAGGCGCTTGGGAGAGACCTGTGGGCGCATCG 2880  
QY 2881 TGGAGCGTGGCGTATCCGCACTCTGCTCGCTGAGAGATATTTCTGTGTG 2940  
DB 2881 TGGAGCGTGGCGTATCCGCACTCTGCTCGCTGAGAGATATTTCTGTGTG 2940  
QY 2941 TAAAGGACATGGGGCAAACTGAAGTGTGATTAATAAAGTTAAACAAACAAAC 3000  
DB 2941 TAAAGGACATGGGGCAAACTGAAGTGTGATTAATAAAGTTAAACAAACAAAC 3000  
QY 3001 AAAAAACAAACAAACAAATTAACCAACAGACGAGAGACGGAAGGTTCAACACC 3060  
DB 3001 AAAAAACAAACAAACAAATTAACCAACAGACGAGAGACGGAAGGTTCAACACC 3060  
QY 3061 CAGAGAGAAAGAAATTTAAAGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3120  
DB 3061 CAGAGAGAAAGAAATTTAAAGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3120  
QY 3121 CTGCAAAAGGTTGGAATCATCTCTGAGTGTTCATGTAACTTCAAGTCTAATCTA 3180  
DB 3121 CTGCAAAAGGTTGGAATCATCTCTGAGTGTTCATGTAACTTCAAGTCTAATCTA 3180  
QY 3181 AAAAGCAAAATAGGCGCCCTCCCTCTCCGCTCTAGAGGAGGAACTTTTGTG 3240  
DB 3181 AAAAGCAAAATAGGCGCCCTCCCTCTCCGCTCTAGAGGAGGAACTTTTGTG 3240  
QY 3241 TTCTACTTTTTCAGAGGGGTTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3300  
DB 3241 TTCTACTTTTTCAGAGGGGTTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3300  
QY 3301 AAAAGAGAGTATTGACGAACAAATAGTAACAAACAAAGTAAAGTGCCTTGGAGAG 3360  
DB 3301 AAAAGAGAGTATTGACGAACAAATAGTAACAAACAAAGTAAAGTGCCTTGGAGAG 3360  
QY 3361 GAAAGGAGAGAGGAGAAATTTCTATAAATACTTAATAATGTTTGTGTGTGTGTGTGT 3420  
DB 3361 GAAAGGAGAGAGGAGAAATTTCTATAAATACTTAATAATGTTTGTGTGTGTGTGTGT 3420  
QY 3421 TTTCTATATATCTCTTGT 3480  
DB 3421 TTTCTATATATCTCTTGT 3480  
QY 3481 AATAGAGACCTCGAGAGGAGAGTCTCTCCACCCCGGAGAGTCTCAACAGACACC 3540  
DB 3481 AATAGAGACCTCGAGAGGAGAGTCTCTCCACCCCGGAGAGTCTCAACAGACACC 3540  
QY 3541 ATTCTGTCTATGCAAAACAGAAACCAACTAGCAGAGGCGCTGAGAGAACCAACACC 3600  
DB 3541 ATTCTGTCTATGCAAAACAGAAACCAACTAGCAGAGGCGCTGAGAGAACCAACACC 3600  
QY 3601 AGACACTTTTCTACAGATTTTCAAGTGTCTTACAGTTTATATATATATATATATATAT 3660  
DB 3601 AGACACTTTTCTACAGATTTTCAAGTGTCTTACAGTTTATATATATATATATATATAT 3660  
QY 3661 TCTAGAAAGCCGCTGTTACCTCTGTTTACAGTTTATATATATATATATATATATATAT 3720  
DB 3661 TCTAGAAAGCCGCTGTTACCTCTGTTTACAGTTTATATATATATATATATATATATAT 3720  
QY 3721 ATATATATTAAGAGTACTGTATTAATCTGTATCATCCGACTTCAATATGTTGCTTTCAAA 3780  
DB 3721 ATATATATTAAGAGTACTGTATTAATCTGTATCATCCGACTTCAATATGTTGCTTTCAAA 3780  
QY 3781 ACAGCAGATGAGCAAAAGACATGAGCTTCCGCTGAGCAAGGTTTTCAGACC 3840  
DB 3781 ACAGCAGATGAGCAAAAGACATGAGCTTCCGCTGAGCAAGGTTTTCAGACC 3840  
QY 3841 CAGGATGGGAGAGGAGACACTGAGAGGGGTTTAAACAACTGAAGATGACCAATAT 3900

Db 3841 CAGGATGGGAGAGGGGAGCAGCTGGAGGGGGTTTAAACAACTGAAGATGACCCATAT 3900  
Qy 3901 CACCCCCACCCCTGCTCCCTAGCTTACCTGCTCAAAAAGGGGCTCAGCTGAGGTG 3960  
Db 3901 CACCCCCACCCCTGCTCCCTAGCTTACCTGCTCAAAAAGGGGCTCAGCTGAGGTG 3960  
Qy 3961 GTCGACCTGGGAGAGCTGAGTGGGAATTTATCCAGACTCGGTGCAATAACCTTGA 4020  
Db 3961 GTCGACCTGGGAGAGCTGAGTGGGAATTTATCCAGACTCGGTGCAATAACCTTGA 4020  
Qy 4021 ATATGAATATAAATGACTGCTCAGAAAAATGCTTGAGAAAAATGCTGCTGATTTT 4080  
Db 4021 ATATGAATATAAATGACTGCTCAGAAAAATGCTTGAGAAAAATGCTGCTGATTTT 4080  
Qy 4081 GAATTCGTGAGCAGCTGAGGAGCCCTTCTGGGATCAGAAATATCCAGAGTGGGGA 4140  
Db 4081 GAATTCGTGAGCAGCTGAGGAGCCCTTCTGGGATCAGAAATATCCAGAGTGGGGA 4140  
Qy 4141 AGTGACCCGCCATTAAACCCNCTGGAGCAAAATAAAAAATCAATAAATGT 4192  
Db 4141 AGTGACCCGCCATTAAACCCNCTGGAGCAAAATAAAAAATCAATAAATGT 4192

## RESULT 2

AAZ37097  
ID AAZ37097 standard; DNA; 4416 BP.  
XX  
AC AAZ37097;  
XX  
DT 27-MAR-2000 (first entry)  
XX  
DE DNA encoding de novo DNA cytosine methyltransferase DNMT3A.  
XX  
KW De novo DNA cytosine methyltransferase; DNMT3A; neoplastic disorder;  
KW carcinoma; sarcoma; leukaemia; DNA methylation; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key  
FT CDS  
FT Location/Qualifiers  
FT 353..391  
FT /\*tag= a  
FT /product= "de novo DNA cytosine methyltransferase  
FT DNMT3A"

WO9967397-A1.

29-DEC-1999.

25-JUN-1999; 99WO-US14373.

25-JUN-1998; 98US-0090906.

24-JUL-1998; 98US-0093993.

(GEHO) GEN HOSPITAL CORP.

Li E, Okano M, Xie S;

WPI: 2000-106298/09.

P-PSDB; AAY54057.

New mouse and human polypeptides, useful to treat and diagnose

neoplastic disorders e.g. carcinomas, sarcomas and leukemias -

Claim 10; Fig 1C; 114pp; English.

The present sequence encodes a human de novo DNA cytosine methyltransferase designated DNMT3A. The polypeptides can be administered therapeutically, especially by expressing encoding polynucleotides, to treat diseases associated with DNA cytosine methyltransferase, such as neoplastic disorders e.g. carcinomas, sarcomas and leukemias. They can be used to diagnose, or determine susceptibility to neoplastic disorders, by assaying for polypeptide expression levels in mammalian cells/body

CC fluids. They are useful to screen for compounds inhibiting/activating the polypeptide. The polypeptides can also be used for in vitro de novo methylation of DNA. Such in vitro methylation may be used to direct or regulate DNA expression in biological systems, e.g. recombinant DNA methylated in vitro may be introduced into a cell/organism to increase or decrease expression of a desired polypeptide for which the native DNA is under-methylated or not methylated. The polypeptides can also be used to produce antibodies which are useful to detect and purify the polypeptide or therapeutically e.g. to treat neoplastic disorders. The polynucleotides are useful to produce probes and primers which are useful diagnostically.

XX.

SQ Sequence 4416 BP; 1107 A; 1131 C; 1327 G; 851 T; 0 other;

Query Match 68.5%; Score 2871.8; DB 21; Length 4416;

Best Local Similarity 83.7%; Pred. No. 0;

Matches 3574; Conservative 0; Mismatches 583; Indels 113; Gaps 24;

Qy 17 CCGGGCGCCCGACCCCGGGCCACACGGCAGAGCCCTGAAAGCCCGAGCCCTGAGGCT 76  
Db 147 CCGGGCGCCCGACCCCGGGCCACACGGCAGAGCCCTGAAAGCCCGAGCCCTGAGGCT 206  
Qy 77 G-----CACTTTTCGAGGGCTTGACATCAGGGCTCTATGTTTAAAGTCTTACTCT 126  
Db 207 GACAGAGGCACCGTTCCACAGAGGGCTCAACACCGGATCTATGTTTAAAGTCTTACTCT 266  
Qy 127 TGCTTAAAGAGACACGGCAATTCCTTCTGAGAGCCCTCGAGCCCGCCACAGCCCTCG 186  
Db 267 CGCTTCAAGAGACACGATATTCCTTCTCCCAAGCCCGC-AGCAGCCCGCCCGC-CCGCG 324  
Qy 187 CAGCCCGAGCTGCGGCTTACTGCCCA-----GCAATGCCCTCCAGCGGCCCGGGGA 239  
Db 325 CAGCCCGAGCTGCGGCTTCCCGGGCCAGATGCCCGCCATGCCCTCCAGCGGCCCGGGGA 384  
Qy 240 CACAGCAGCTCTCTCTGAGCGGGAGGATGATGAAAGGAGAGAGAGAGAGAGAGGA 299  
Db 385 CACAGCAGCTCTCTCTGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 444  
Qy 300 GAACCTGGCAAGAGAGAGCCAGAGCCAGAGCCAGCCCGCCGAGAGGAGGAGGAGGCC 359  
Db 445 GCGCGCTGGCAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGCC 504  
Qy 360 TGCGCCGAGCGCAAGCAGCCAGCGGTGGAAGAGTCAACCCCGCCAGAGGAGGAGGAGGCC 419  
Db 505 TGGAGGAGCGCAAGCAGCCAGCGGTGGAAGAGGAGTCAACCCCGCCAGAGGAGGAGGCC 564  
Qy 420 GACCAACCAAGTCTCAGCCCATGCGCCAGGACTCTGGCCCTCAGATCTGCTACCCCAATGG 479  
Db 565 GATCTCAAGTCTCCCATCCATGCGCCAGGACTCAGCGCCCTCAGAGCTATTACCCCAATGG 624  
Qy 480 AGACTTGGAGAAGCGGAGTGAACCCCAACCTGAGAGGGGAGCCCGAGCTGAGGGGAGAA 539  
Db 625 GGACTTGGAGAGCGGAGTGAACCCCAACCTGAGAGGGGAGGAGGAGGAGGAGGAGGAG 684  
Qy 540 GGGTGGGGCCCGAGCTGAAGAGAGGG-----AACTGAGAGCCCGCCAGAGAGCTTCCAGAGC 596  
Db 685 GGGCGGGCCCGAGCAGAGAGGAGGGGTGACAGCTCAGAGCCCTGCTGAGAGCTCAAGAGC 744  
Qy 597 TGTGAGAATGGCTGCTGTGTGACCAAGAGAGGCGGTGAGAGCTCTGAGAGAGAGGGCAA 656  
Db 745 AGTGGAAATGGCTGTGTGACCCCAAGAGAGGCGGAGAGGAGGAGGAGGAGGAGGAGG 804  
Qy 657 AGAACAGAGCAGACCAACATCGAATCGAATAATGAGAGGCTCCCGGGGGGAGTGGCG 716  
Db 805 AGAACAGAGGAGACCAACATCGAATCGAATAATGAGAGGCTCCCGGGGGGAGTGGCG 864  
Qy 717 AGGTGGCTTGGGGTGGAGTCCAGCCTCCGTGAGCGCCCATGCGCAAGACTCACCTTCCA 776  
Db 865 GGGTGGCTTGGGGTGGAGTCCAGCCTCCGTGAGCGGCCCATGCGCGAGGCTCACCTTCCA 924  
Qy 777 GGCAGGGGACCCCTACTACATCAGCAACCGAAGGAGTGAAGTGGCTGGCAGCTTGGAA 836  
Db 925 GGCAGGGGACCCCTACTACTACATCAGCAACCGGAGGAGGAGTGGCTGGCAGCTGGAA 984

QY	837	AAAGGAGGCTGAGAAAGAAAGCCAAAGTAAATTGCACTAATGAATGCTGTGAAAGAACCA	896
Db	985	AAGGGAGGCTGAGAAAGAAAGCCAAAGTCAATTGCAGGAATGAATGCTGTGAAAGAAACCA	1044
QY	897	GGCCTTGGAGAGTCTCAGAGGTGGAGAGGCCAGCCCTCTGCTGTGCAGCAGCCAC	956
Db	1045	GGGGCCCGGGAGTCTCAACAGGTGGAGAGGCCAGCCCTCTGCTGTGCAGCAGCCAC	1104
QY	957	GGACCTGCTCTTCTCCGACTGTGGCCACACCCCTGAGCCAGTAGAGAGGGAGTCTGGGA	1016
Db	1105	TGACCCCGCATCCCCCACTGTGGCTACCAAGCTGAGCCCTGGGGTCCGATGCTGGGA	1164
QY	1017	CAAGAAATGCTAACAAAGCAGCCGACGATGAGCCTGAGTATGAGGATGGCCGGGCTTTGG	1076
Db	1165	CAAGAAATGCAACCAAGCAGCGCATGACGACCAAGTACGAGGACGGCCGGGGCTTTGG	1224
Db	1077	CATTGAGAGCTGTGTGTGGGGGAACTTGGGGCTTCTCCTGTGTGGCCAGGCAGAAATGT	1136
Db	1225	CATTGGGAGCTGTGTGTGGGGGAACTGTGGGGCTTCTCCTGTGTGGCCAGGCAGAAATGT	1284
QY	1137	GTCCTGTGTGATGACAGGCGGAGCCGAGCAGCTGAGGCACTCGCTGGTCAATGTGTT	1196
Db	1285	GTCCTGTGTGATGACAGGCGGAGCCGAGCAGCTGAGGCACTCGCTGGTCAATGTGTT	1344
QY	1197	CGAGATGGCAAGTCTCAGTGTGTGTGTGGAGAACTCATGCGCTGAGCTCTCTTG	1256
Db	1345	CGAGATGGCAAGTCTCAGTGTGTGTGTGGAGAACTCATGCGCTGAGCTCTCTTG	1404
QY	1257	CAGTGATTCACACGAGCCACCTACAAACAAGCAGCCATGTACCGCAAGCCATCTACGA	1316
Db	1405	CAGTGATTCACACGAGCCACCTACAAACAAGCAGCCATGTACCGCAAGCCATCTACGA	1464
QY	1317	AGTCCTCAGGTGGCAGCAGCCGCTGCGGAGAGCTGTTTCCAGCTTGCCATGACAGTGA	1376
Db	1465	AGTCCTCAGGTGGCAGCAGCCGCTGCGGAGAGCTGTTTCCAGCTTGCCATGACAGTGA	1524
QY	1377	TGAAAGTGAACAGTGGCAAGGCTGTGGAAAGTGCAGAAACAAGCAGATGATTTGAATGGGCCCT	1436
Db	1525	TGAGAGTGAACATGCGCAAGCCGCTGGAGGTGCAGAAACAAGCCATGATTTGAATGGGCCCT	1584
QY	1437	CGGTGGCTTCCAGCCCTCGGGTCTTAAGGGCCTGGAGCCACACGAAGAGAGAAAGATCC	1496
Db	1585	CGGGGGCTTCCAGCATTTATGGCCCTTAAGGGCCTGGAGCCACACGAAGAGAGAAAGATCC	1644
QY	1497	TTACAAGGAAGTTTACACCGACATGTGGGTGGAGCCTGAGAAAGAGCTGCTTACGCCACAC	1556
Db	1645	CTACAAAGAGTGTACACCGACATGTGGGTGGAACTTGAGCAGCTGATGACACACAC	1704
QY	1557	CCACACGAGCAAGAAACCAGAAAGACACAACAGAAAGAACTTAAGGTCAAGGAGATCAT	1616
Db	1705	TCACACGAGCAAAAGAGCCCGGAGAGACACAGCGAGAAAGCCAAAGTCAAGAGAGATTA	1764
QY	1617	TGATGAGCGCACAAAGGAGCGGCTGTGTATGAGAGTGGCCACAGAAAGTGCAGAAACATCGA	1676
Db	1765	TGATGAGCGCACAAAGAGCGGCTGTGTATGAGAGTGGCCAGAAAGTGCAGAAACATCGA	1824
QY	1677	GGACATTTGTATCTCATGTGGGAGCCTCAATGTCAACCTGGAGACCCACTTTCATTGG	1736
Db	1825	GGACATTTGTATCTCTGTGGAGCCTCAATGTTAACCTGGAAACACCCCTTTCGTTGG	1884
QY	1737	AGGCATGTGCAGAACTGTAAAGACTGCTTTGGAGTGTGCTTACAGTATGACGACGA	1796
Db	1885	AGGCATGTGCAGAACTGTAAAGACTGCTTCTGGAAGTGTGCTTACAGTATGACGACGA	1944
QY	1797	TGGGTACCAAGTCTCAATGACACATGCTGTGGGGGGCGGAAAGTGCATATGTGTGGAA	1856
Db	1945	CGGCTACCAAGTCTCAATGACACATGCTGTGGGGGGCGGAAAGTGCATATGTGTGGAA	2004
QY	1857	CAACAACTGTGCAGGTGCTTTTGTGTGTGCAGATGTGTGATCTCTTGGTGGGGCCAGGAGC	1916
Db	2005	CAACAACTGTGCAGGTGCTTTTGTGTGTGTGCAGATGTGTGATCTCTTGGTGGGGCCAGGAGC	2064

Qy	1917	TGCTCAGGCAAGCCATTAAAGAAAGACCCCTGGAACCTGCTACATGTGCGGGCATAAAGGCAC	1976
Db	2065	TGCCCCAGGCAAGCCATTAAAGAAAGACCCCTGGAACCTGCTACATGTGCGGGCAAAAGGCTAC	2124
Qy	1977	CTATGGGCTGCTGCGAAGACGGGAAGAAGCTGGCCCTTCTGAGCTCAGATGTTTCTTTGGCAA	2036
Db	2125	CTACGGGCTGTGCGCGGGCGGAAGAGACTGGCCCTCCGGGCTCCAGATGTTCTTGCTTAA	2184
Qy	2037	TAACCATGACCAAGAAATTTGACCCCGCCAAAAGTTTACCACCTGTGCAAGCTGAGAGAG	2096
Db	2185	TAACCAAGCAAGAAATTTGACCCCTCCAAAAGSTTTAACCAACCTGTCCACGCTGAGAAAAG	2244
Qy	2097	GAAGCCATTCGGGTGCTGTCTCTCTTTGATGGGAATTGTCTACAGGCTCCTGTGCTGAA	2156
Db	2245	GAAGCCATTCGGGTGCTGTCTCTCTTTGATGGGAATGCTACAGGGCTCCTGTGCTGAA	2304
Qy	2157	GGACCTGGGCATCCAAGTGAAGCCGCTACATTTGCTCCGAGGTGTGTAGAGACTTCATCAC	2216
Db	2305	GGACTTGGGCAATTCAGGTGAAGCCGCTACATTTGCTCCGAGGTGTGTAGAGACTTCATCAC	2364
Qy	2217	GGTGGGCATGTGCGGCAACCAAGGAAAGATCATGTACGTGGGGGACGTCCGCAAGCTCAC	2276
Db	2365	GGTGGGCATGTGCGGCAACCAAGGAAAGATCATGTACGTGGGGGACGTCCGCAAGCTCAC	2424
Qy	2277	ACAGAAACATATCCAGAGAGTGGGGCCCATTCGACCTGTGATTGGAGGCAAGTCCCTGCAA	2336
Db	2425	ACAGAAACATATCCAGAGAGTGGGGCCCATTCGACCTGTGATTGGAGGCAAGTCCCTGCAA	2484
Qy	2337	TGACCTTCCATTGTCAAACCTTGCCCGCAAGGACCTTTATGAGGGTACTGGCCGCTCTT	2396
Db	2485	TGACCTTCCATTGTCAAACCTTGCCCGCAAGGACCTTTATGAGGGTACTGGCCGCTCTT	2544
Qy	2397	CTTTGAGTTCTACCGGCTCTTGCAATGATGCGGGCCCAAGAGAGAGATGATCGGCCCTT	2456
Db	2545	CTTTGAGTTCTACCGGCTCTTGCAATGATGCGGGCCCAAGAGAGAGATGATCGGCCCTT	2604
Qy	2457	CTTCTGGCTCTTTGAGAAATGTGTGGCCATGCGCGTTAAGTGAACAAGGGAACATCTCGCG	2516
Db	2605	CTTCTGGCTCTTTGAGAAATGTGTGGCCATGCGCGTTAAGTGAACAAGGGAACATCTCGCG	2664
Qy	2517	ATTCTTGAAGTCTAACCCCGGTGATGATTGAACGCAAGAAGTGTCTGTGCAACAGAGGC	2576
Db	2665	ATTCTTGAAGTCTAACCCCGGTGATGATTGAATGCCAAGAAGTGTCACTGCAACAGAGGC	2724
Qy	2577	CCGTTACTTCTGGGGTAAACCTTCTGAGCATGAACAGGCTTTTGCACTCACTGTGAATGA	2636
Db	2725	CCGTTACTTCTGGGGTAAACCTTCCGSGTATGAACAGGCGCTTTGGCATCACTGTGAATGA	2784
Qy	2637	TAAGCTGAGCTGCAGAAGTGTCTGAGACGCGGAGAAATAGCCMAGTTCAGCAAAAGTGA	2696
Db	2785	TAAGCTGAGCTGCAGAAGTGTCTGAGACATGGCAGATAGCCMAAGTTCAGCAAAAGTGA	2844
Qy	2697	GACCAATTACCAACAGGTCAAACTCTATAAAGCAGGCGAAAGACCAGCAATTTCCCGTCTT	2756
Db	2845	GACCAATTACTACAGGTCTAACTCCATTAAGCAGGCGAAAGACCAGCAATTTCCGTCTT	2904
Qy	2757	CATGAACGAGAGAGAGACATCTCTGTGTGACACTGAATAAGAAAGGTGTTTGCTTCCC	2816
Db	2905	CATGAATGAGAAAGAGACATCTTATGTGTGCACTGMAATGMAAAGGTAATTTGGTTCCC	2964
Qy	2817	CGTCCACTACACAGACGTCTCCAACTGAGCCGCTTGCGAGGCGAGACGACTGTGGCGCG	2876
Db	2965	AGTCCACTATACTGACGTCTCCAACTGAGCCGCTTGCGAGGCGAGACGACTGTGGCGCG	3024
Qy	2877	ATCGTGAAGCGTCCGGTCACTCCGCCACCTTTCGCTCCGCTGMAAGAAATATTTTGCTTG	2936
Db	3025	GTCAATGAGCGTCCGAGTCACTCCGCCACCTTTCGCTCCGCTGMAAGAAATATTTTGCGTG	3084
Qy	2937	TGTGTAAAGGACATGGGGGCAAACTGAAGTATGATGATTAATAAAAGTTTAAACAACAAC	2996
Db	3085	TGTGTAAAGGACATGGGGGCAAACTGAAGTATGATGATTAATAAAAGTTTAAACAACAAC	3140
Qy	2997	AAACAAAAAACAACCAATATAAACAACAAGAACGAGAGACGAGAAAAAGT-TCA	3055

Db	3141	-----AAAAAACACAAAACATAAATAAAACACCAAGAACATGAGGATGGAGAGAAGTATCA	3195
Qy	3056	GCACCCAGAGAGAAAAGGAATTTAAAGC--AAACACACAGAGGAGGAAAACGCCGGAGG	3113
Db	3196	GCACCCAGAGAGAAAAGGAATTTAAACAAAACACACAGAGCGGAAATATACCGAGGG	3255
Qy	3114	GCATTGGCCTTGCAAAAAGGGTTGGACATCATCTCTCGAGTTTTCAATGTTTAACTTCAGTTC	3173
Db	3256	CTTTGGCCTTGCAAAAAGGGTTGGACATCATCTCTCGATTTTTCAATGTATATTTCTTCAGTTC	3315
Qy	3174	CTATCTAAAAAGCAAAATPAGGC-CCCTCCCTTCTTCCCTC---CCGGTCTTAGAGAGCG	3229
Db	3316	CTATTTAAAAACAACAAACAGCTCCCTTCCCTTCCCTCCCTTCTTCTTTTCGGTCA	3375
Qy	3230	AACATTTTGTGTTTTCTACTCTTTTTTCAGAGGGTTTTCTGTGTTTTGGTTTTGTTTCT	3289
Db	3376	GACCTTTTATTTTCTACTCTTTT-CAGAGGGTTTTCTGTGTTGTTTGGG-TTTTGTGTTCT	3433
Qy	3290	TGCTGTGACTCAAAACAAGAGAGTTATTGCGAGC-AAAATCAGTAACAAACAAAAAGTAGAAA	3348
Db	3434	TGCTGTGACTGAACAAGAGAGTTATTGCGACAAAATCAGTAACAAAAATAGTAACAA	3493
Qy	3349	TGCTTGAGAGAGAAAG-----GGAGAGAGGGAATTTCTATATAAAACTTAAAA	3397
Db	3494	TACCTTGAGAGAAAGGTGGGAGGAGAGAAAAGGAAATTTTTAAAGAAATCTATA	3553
Qy	3398	TATTGGTTTTTTTTTTTTTTTTTTCCTT-----TC	3424
Db	3554	TATTGGGTGTTTTTTTTTTGTTTTTTTTTGGTTTTTTTTTTTTTTTACTA	3613
Qy	3425	TATATATCTCTTTGGTTGTTCTCTAGCTGATCAGATAGGAGACAAACA--CGAAGAGAA	3482
Db	3614	TATATCTTTTTTTTTTGTGTTCTCTAGCTGATCAGATAGGAGACCAAGCAGGGGACGAAA	3673
Qy	3483	TAGAGACCCCTCGGAGGAG-AGTCTCTCTCTCCACCCCGCAGCAGTCTCAACAGCACCA	3541
Db	3674	GAGAGAGACATCAGCGCGCAGCATTCCTCCCGACCACTGAGCTGCTGCGCAGCACCA	3733
Qy	3542	TTCTGTTGATGCAAAAACAGAACCAACTAGCAGCAGGGCGCTGAGAGAACACCAACCA	3601
Db	3734	TTCTGTTGACGCAAAAACAGAACCCAGTTAGCAGCAGGGAGACAGAACCAACACACGA	3793
Qy	3602	GACACTTTCTACAGTATTTTCAGGTGCTTACCACACAGGAAACCTTGAAGAAAACCAAGTTT	3661
Db	3794	CA-TTTTTTCTACAGTATTTTCAGGTGCTTACCACACAGGAAACCTTGAAGAAATCAGTTT	3852
Qy	3662	CTAGAGCGGCTGTTACCTCTGTTTTCAGTTTATATATATATATAGATAGATATAGATATA	3721
Db	3853	CTAGAAGCGGCTGTTACCTCTGTTTTCAGTTTATATATATATATATAGATAGATAGAG--ATA	3910
Qy	3722	TATATATAAAGGTACTGTTAACTACTGTATCCCGACTTTCATATATGGTGTCTTCAAAA	3781
Db	3911	TATATATAAAGGTACTGTTAACTACTGTATCAACCGACTTTCATATATGGTGTCTTCAAA	3969
Qy	3782	CAGCGAGATGACAAAGACATCAGCTTCGCTCGGCCCTCTGTGCAAAAGGGTTTCAGCCC	3841
Db	3970	CAGCGAGATGAGTAAAAAACATCAGCTTCCAGTTGCTCTTCGCAAAAGGGTTTCACCAA	4029
Qy	3842	AGGATGGGGAGGGGAGCAGCTGGAGGGGGTTTTAACAAACTGAAGAGATGACCCATATC	3901
Db	4030	GGATGGAAAGGGAGACAGCTTGACATGGGCGGTTCTCAGGTGGGCTCTTCTCCCTTGG	4089
Qy	3902	ACCCCCAC-----CCCTGCCCATGCTAGCTTCCACCTGCCCAAAAGGGGC	3948
Db	4090	GTTTGTAACGAAGTGAAGGAGAGAACTTGGGAGCCAGGTTCTCCCTGCTCCAAAAGGGGG	4149
Qy	3949	TCAGCTGAGGTGGTCGACCCCTGGGGAAGCTGAGTGTGAAATTTATCCAGACTGCGGTGC	4008
Db	4150	CTAGATGAGGTGGTCGGGCCCTGGACAGCTGAGAGTGGGATTCATCCAGACTCATGCAA	4209
Qy	4009	AATAACCTTAGAATATGAATCTAAAAATGACTGCTCAGAAAAAATGSC-----TTGAGAA	4062

Db	4210	TAA	CCCTTTGATTGTTTCTTAAAGGAGACATCCCTCGGCAAGATGGCGAGAGGGTACGGAG	4269
Qy	4063	AAC	ATTGCTCCCTGATTTTGAATTCGTGAGCCACGTTGAAGGCCCTTCTGTGGATCAGAAA	4122
Db	4270	TC	TCAGGCCAGTTTCTCACTTTAGCCAAATTCGAGGGCTCCTTGTGGTGGATCAGAAC	4329
Qy	4123	TAT	TCCAGATGAGGGAAAGTGACCCGCCATTACCCCNCCCTGGAGCAATAAATAAAACA	4182
Db	4330	TAA	TCCAGAGTGTGGGAAAGTGACAGTCAA--AAACCCCACTGGAGCAATAAATAAAACA	4387
Qy	4183	TACA	AAATGT 4192	
Db	4388	TACA	AAACGT 4397	
RESULT 3				
ABL90391				
ID ABL90391 standard; cDNA; 2938 BP.				
XX	AC	ABL90391;		
XX	DT	24-MAY-2002 (first entry)		
XX	DE	Human polynucleotide SEQ ID NO 953.		
XX	KW	Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;		
XX	KW	antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;		
XX	KW	vulnerable; anticonvulsant; antibacterial; antifungal; antiparasitic;		
XX	KW	cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;		
XX	KW	neurological disease; infection; human; secreted protein; gene; ss.		
OS		Homo sapiens.		
XX	PN	WO200190304-A2.		
XX	PD	29-NOV-2001.		
XX	PF	18-MAY-2001; 2001WO-US16450.		
XX	PR	19-MAY-2000; 2000US-205515P.		
XX	PA	(HUMA-) HUMAN GENOME SCI INC.		
XX	PI	Birse CE, Rosen CA;		
XX	DR	WPI; 2002-122018/16.		
XX	DR	P-PSDB; ABB89982.		
XX	PT	Novel 1405 isolated polypeptides, useful for diagnosis, treatment and		
PT	PT	prevention of neural, immune system, muscular, reproductive,		
PT	PT	gastrointestinal, pulmonary, cardiovascular, renal and proliferative		
XX	PS	disorders -		
XX	PS	Claim 4: SEQ ID NO 953; 2081bp + Sequence Listing; English.		

The invention relates to novel genes (ABL89449-ABJ90853) and proteins (ABB93040-ABB90444) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune; (c) haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (d) cardiovascular disorders such as myocardial ischaemias; (e) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published/pct](http://ftp.wipo.int/pub/published/pct) sequences.

XX Sequence 2938 BP; 759 A; 695 C; 774 G; 700 T; 10 other;

Query Match 43.3%; Score 1814; DB 24; Length 2938;

Best Local Similarity 82.5%; Pred.No. 0;

Matches 2327; Conservative 5; Mismatches 398; Indels 89; Gaps 19;

QY 1446 CCAAGCCCTGGGCTCTTAAGGAGCTGGAGCCACAGAGAGAGAGAAATCCTTACAGGA 1505  
DB 9 CCAAGCCCTGGGCTCTTAAGGAGCTGGAGCCACAGAGAGAGAGAAATCCTTACAGGA 68  
QY 1506 AGTTTACACCAATGTGGGTGAGAGCTTGAAGAGCTGCTTACGCCCCACCCACCAGC 1565  
DB 69 AGTGTACACGACATGTGGGTGAGAGCTTGAAGAGCTGCTTACGCCCCACCCACCAGC 128  
QY 1566 CAAAGAAACCCAGAAAGAGCAACAAGAGAACTTAAAGTCAAGAGATCATTTGATGAGCG 1625  
DB 129 CAAAGAAACCCAGAAAGAGCAACAAGAGAACTTAAAGTCAAGAGATCATTTGATGAGCG 188  
QY 1626 CACAGAGAGCGGCTGTGTATGAGGTGGCCAGAAAGTGCAGAAACATTCAGAGACATTTG 1685  
DB 189 CACAGAGAGCGGCTGTGTATGAGGTGGCCAGAAAGTGCAGAAACATTCAGAGACATTTG 248  
QY 1686 TATCTCATGTGGAGGCTCTCATGTCAACCTTGAGCAACCACTTTTCAATTGAGGATGTG 1745  
DB 249 CATCTCCGTGGAGGCTCTCATGTCAACCTTGAGCAACCACTTTTCTGTGAGGAAATGTG 308  
QY 1746 CCAGAACCTGTAAAGATCTCTTGTGAGGTGTGCTTACAGATATGACAGATGGATACCA 1805  
DB 309 CCAAAACCTGTAAAGATCTCTTGTGAGGTGTGCTTACAGATATGACAGATGGATACCA 368  
QY 1806 GTCTTATGACACATGTCTGTGGGGGCGTGAAGTCTCATGTGTGGAAACAACAATG 1865  
DB 369 GTCTTATGACACATGTCTGTGGGGGCGTGAAGTCTCATGTGTGGAAACAACAATG 428  
QY 1866 CTGAGAGGCTTTTGTGTGAGGTGTGATCTCTTGTGTGGGGCCAGAGCTCTCAAGC 1925  
DB 429 CTGAGAGGCTTTTGTGTGAGGTGTGATCTCTTGTGTGGGGCCAGAGCTCTCAAGC 488  
QY 1926 AGCCATTAGAGAAAGCCCTGGAACTGCTACATGTGCGGGCATTAAGGCACTATGGGCT 1985  
DB 489 AGCCATTAGAGAAAGCCCTGGAACTGCTACATGTGCGGGCATTAAGGCACTATGGGCT 548  
QY 1986 GCTGCGAAGACGGGAAGACTGGCTTCTGCACTCCAGATGTTCTTGGCAATACCATGA 2045  
DB 549 GCTGCGGCGGCGAGAGAGACTGGCTTCTGCACTCCAGATGTTCTTGGCAATACCATGA 608  
QY 2046 CCAGGAATTTGACCCCCAAAGTTTACCACCTGTGCGAGCTGAGAGAGAGAGCCCAT 2105  
DB 609 CCAGGAATTTGACCCCCAAAGTTTACCACCTGTGCGAGCTGAGAGAGAGAGCCCAT 668  
QY 2106 CCGGCTGTGCTCTCTTTGATGAGGATGCTACAGGGCTCTGTGCTGAGAACCTGGG 2165  
DB 669 CCGGCTGTGCTCTCTTTGATGAGGATGCTACAGGGCTCTGTGCTGAGAACCTGGG 728  
QY 2166 CATTCAAGTGAACCGCTACATTTGCTCCGAGGTGTGAGAGATTCATACGTTGGGCT 2225  
DB 729 CATTCAAGTGAACCGCTACATTTGCTCCGAGGTGTGAGAGATTCATACGTTGGGCT 788  
QY 2226 GGTGCGGACCAAGGAAAGATCATGTAGCTGGGGGAGCTCCGCGAGGCTCACACAGAGCA 2285  
DB 789 GGTGCGGACCAAGGAAAGATCATGTAGCTGGGGGAGCTCCGCGAGGCTCACACAGAGCA 848  
QY 2286 TATCAGAGTGGGGCCCATTCGACCTGTGATTTGAGGACAGTCCCTGAGATGACCTCTC 2345  
DB 849 TATCAGAGTGGGGCCCATTCGACCTGTGATTTGAGGACAGTCCCTGAGATGACCTCTC 908  
QY 2346 CATGTCAACCTGCGCGAGAGGACTTATGAGGATGAGCGGCTCTTCTTGATG 2405  
DB 909 CATGTCAACCTGCTGCGAGAGGCTCTTACAGAGGACTGGCGGCTCTTCTTGATG 968  
QY 2406 CTACCGCTCTGATGATGAGCGGCGCAAGAGAGAGATGAGCCCTTCTTGAGCT 2465

DB 969 CTACCGCTCTGATGATGCGGCGCCCAAGAGAGAGATGATGCCCTTCTTGAGCT 1028  
QY 2466 CTTTGAAGATGTGTGGCCATGAGGCGCTTATGTACAAAGAGGACATCTGCGATTTCTTGA 2525  
DB 1029 CTTTGAAGATGTGTGGCCATGAGGCGCTTATGTACAAAGAGGACATCTGCGATTTCTGCA 1088  
QY 2526 GTCTAACCCCGTATGATGATGAGCCCAAGAAAGTGTCTGTGCACACAGGCGCTTACTT 2585  
DB 1089 GTCCAACCCGTGATGATGATGATGCCAAAGAGTGTGAGGTGCACACAGGCGCTTACTT 1148  
QY 2586 CTGGGGTAACTTCTGCGATGAAACAGGCGCTTGTGATCCATGTAATGATAGCTGGA 2645  
DB 1149 CTGGGGTAACTTCTGCGATGAAACAGGCGCTTGTGATCCATGTAATGATAGCTGGA 1208  
QY 2646 GCTGCAAGAGTGTCTGAGACAGGCGAGATATGACCAAGTTTCAGCAAAAGTGAAGCACTTAC 2705  
DB 1209 GCTGCAAGAGTGTCTGAGACAGGCGAGATATGACCAAGTTTCAGCAAAAGTGAAGCACTTAC 1268  
QY 2706 CACCAAGTCAAACTCTATTAAGAGGCGCAAGAGCAATTTCCCGTCTTATGACGA 2765  
DB 1269 TACGAGGTCAAACTCTATTAAGAGGCGCAAGAGCAATTTCCCGTCTTATGACGA 1328  
QY 2766 GAAAGAGACATCTCTGTGTGACCTGAATGAAAGGCTTTGGCTTCCCGTCCACTTA 2825  
DB 1329 GAAAGAGACATCTCTGTGTGACCTGAATGAAAGGCTTTGGCTTCCCGTCCACTTA 1388  
QY 2826 CACAGAGCTCTTCAACATGAGCGGCTTGGCGAGAGAGACTGTGGGCGCATGCGAG 2885  
DB 1389 TACTGAGCTCTTCAACATGAGCGGCTTGGCGAGAGAGACTGTGGGCGCATGCGAG 1448  
QY 2886 CGTGCCGCTCATCCGCGACCTCTTCCGCTGGAAGAAATTTTGTGTGTGAAG 2945  
DB 1449 CGTGCCGCTCATCCGCGACCTCTTCCGCTGGAAGAAATTTTGTGTGTGAAG 1508  
QY 2946 GACATGGGGGGAATGAGTGTGATGATTAAGAAAGTTTAAACAACAACAACA 3005  
DB 1509 GACATGGGGGGAATGAGTGTGATGATTAAGAAAGTTTAAACAACAACAACA 1559  
QY 3006 ACAAAACAACAATTAATAACCAAGAGAGAGAGAGAGAGAGTCTCAGCACCCAGA 3064  
DB 1560 ACAAAACAACAATTAATAACCAAGAGAGAGAGAGAGAGTCTCAGCACCCAGA 1619  
QY 3065 AGAGAAAGAGATTTAAAGC-AAACACAGAGAGAGAGAGAGAGAGGCTTGGCT 3122  
DB 1620 AGAGAAAGAGATTTAAAGC-AAACACAGAGAGAGAGAGAGAGAGGCTTGGCT 1679  
QY 3123 TCGAAAGGTTGACATCATCTCCGAGTTTCAATGTTAACTTACCTTACTATAA 3182  
DB 1680 TCGAAAGGTTGACATCATCTCCGAGTTTCAATGTTAACTTACCTTACTATAA 1739  
QY 3183 AAGCAAAATAGGC-CCCTCCCTTCTTCCCT--TCCGCTCAGAGGCGAACTTTTG 3238  
DB 1740 AAGCAAAATAGGC-CCCTCCCTTCTTCCCT--TCCGCTCAGAGGCGAACTTTTG 1799  
QY 3239 TTTTCTACTCTTTTCAAGGGGCTTCTGTTGTTGGTTTGTCTGCTGAGC 3288  
DB 1800 TTTTCTACTC-TTTTCAAGGGGCTTCTGTTGTTGGG-TTTTGTCTGCTGAGC 1857  
QY 3299 TGAACAGAGAGATTAATTCAGC-AAATCAGTAACAACAAGAGTGAATGCTTGA 3357  
DB 1858 TGAACAGAGAGATTAATTCAGC-AAATCAGTAACAACAAGAGTGAATGCTTGA 1917  
QY 3358 GAGGAAAGG-----GAGAGAGGAAATTTATTAATAAATTAATAATGCTTGA 3408  
DB 1918 GAGGAAAGGTTGGAGAGAGAGAGAGAGAAATTTATTAATAATTAATAATGCTTGA 1977  
QY 3409 TTTTCTTCTCTT-----TCTATATATCTCT 3435  
DB 1978 TTTTCTTCTCTT-----TCTATATATCTCT 2037  
QY 3436 TTGTTGCTCTAGCCTGATCAGATAGAGACAAACA--GGAAGAGATAGAGACCTC 3493  
DB 2038 TTGTTGCTCTAGCCTGATCAGATAGAGACAAACAAGAGAGAGAGAGAGACAC 2097



```
QY 3494 GGAGGAG-AGTCTCTCTCCACCCCGAGCGAGTCTCAACAGACACCATCTCTGGTGTAT 3552
DB 2098 TCAGCGGGACATTTCTCCCTCCAGCCACTGAGCTGTGTGCGCAGCACCATTCTGTGTCTAC 2157
QY 3553 GCAAAACAGAACCACTAGCAGCAGGGCGTGAAGAACACCAACACAGACACATTTCTTA 3612
DB 2158 GCACAAACAGAACCCAGTTAGCAGCAGGGAGAGACACACACAGACA-TTTTCTTA 2216
QY 3613 CAGTATTTTCAGTGTCTTACACACAGAGAACCTTTGAAGAACACAGTTTCTAGAGCCGC 3672
DB 2217 CAGTATTTTCAGTGTCTTACACACAGAGAACCTTTGAAGAACATCAGTTTCTAGAGCCGC 2276
QY 3673 TGTACCTCTCTGTTCAGTGTATATATATATATATATATATATATATATATATATATAA 3732
DB 2277 TGTACCTCTCTGTTCAGTGTATATATATATATATATATATATATATATATATATATAA 2334
QY 3733 GGTACTGTTAACTACTGCTACATCCGACTTTCATATGCTGCTTTCAAAACAGCGAGTGA 3792
DB 2335 GGTACTGTTAACTACTGCTACATCCGACTTTCATATGCTGCTTTTCAAAACAGCGAGTGA 2393
QY 3793 GCAAAAGACATCAGCTTCCGCTCTGCTCTGTGCAAGGGTTTCAGCCAGGAGGGGAG 3852
DB 2394 GTAAAAACATCAGCTTCCGCTCTGCTGCGCAAGGGTTTCACCAAGGATGAGAAA 2453
QY 3853 AGGGGAGCAGCTGGAGGGGTTTAAACAACCTGAAGGATGACCATATCACCCCCAC-- 3910
DB 2454 GGGAGACAGCTTCAGATGGCGTTCTCAGGTGGGCTCTTCCCTTGGTTTGAACA 2513
QY 3911 -----CCCTGCCCATGCTAGCTTCACTGCTGCAAAAGGGGCTCAGCTGAGGT 3959
DB 2514 AGTGAAGGAGGAGAACTTGGGAGCAGGTCTTCCCTGCGCAAAAGGGGCTAGATGAGT 2573
QY 3960 GGTGCGACCTGCGGAGAGCTGAGTGTGGAATTTATCCAGACTCGCGTGCATTAACCTTAG 4019
DB 2574 GGTGCGGCGGTGGACAGCTGAGAGTGGGATTCATCCAGACTCATGCAATACCTTTGA 2633
QY 4020 AATATGAATCTAAATGAGCTGCTCAGAAAATGCG-----TTGAGAAACATTTGTC 4073
DB 2634 TTGTTTCTTAAAGGAGACTCCCTCGCAAGATGGCAGGGTACGGAGTCTTCAGGGCCC 2693
QY 4074 TGATTTTGAATTCGTGAGCCAGCTTGAAGGCCCTTGTGGATCAGAAATATTTCCAGAGT 4133
DB 2694 AGTTTCTCACTTTAGCCAAATTCGAGGGCTCTTGTGTGGATCAGAACTAATCCAGAGT 2753
QY 4134 GAGGGAAGTGCCCGCATTAACCCCNCTGGAGCAATATAAAACATACAAATGT 4192
DB 2754 GTGGGAAGTGACAGTCAA--AACCCCACTGGAGCAATATAAAACATACAAACGT 2810

RESULT 4
AAZ37096
ID AAZ37096 standard; DNA; 4195 BP.
XX
AC AAZ37096;
XX
DT 27-MAR-2000 (first entry)
XX
DE DNA encoding de novo DNA cytosine methyltransferase Dnmt3b1.
XX
KW De novo DNA cytosine methyltransferase; Dnmt3b1; neoplastic disorder;
KW carcinoma; sarcoma; leukaemia; DNA methylation; ss.
XX
OS Mus sp.
XX
FH Location/Qualifiers
FT 269...2848
CDS /*tag= a
FT /product= "de novo DNA cytosine methyltransferase
FT Dnmt3b1"
XX
EN W09567397-A1.
XX
```

```
PD 29-DEC-1999.
XX 25-JUN-1999; 99NO-US14373.
XX 25-JUN-1998; 98US-0090906.
PR 24-JUL-1998; 98US-0093993.
XX (GEO) GEN HOSPITAL CORP.
XX
PI Li E, Okano M, Xie S;
XX WPI; 2000-106298/09.
DR P-PSDB; AAY54056.
XX
PT New mouse and human polypeptides, useful to treat and diagnose
PT neoplastic disorders e.g. carcinomas, sarcomas and leukemias -
XX
PS Claim 9; Fig 1B; 114pp; English.
XX
CC The present sequence encodes a murine de novo DNA cytosine
CC methyltransferase designated Dnmt3b1. The Dnmt3b gene also produces,
CC through alternate splicing, at least two shorter isoforms of 840 and
CC 777 amino acid residues, termed Dnmt3b2 and Dnmt3b3 (sequences
CC not given in the specification). The polypeptides can be administered
CC therapeutically, especially by expressing encoding polynucleotides, to
CC treat diseases associated with DNA cytosine methyltransferase, such as
CC neoplastic disorders e.g. carcinomas, sarcomas and leukemias. They can
CC be used to diagnose, or determine susceptibility to neoplastic disorders,
CC by assaying for polypeptide expression levels in mammalian cells/body
CC fluids. They are useful to screen for compounds inhibiting/activating the
CC polypeptide. The polypeptides can also be used for in vitro de novo
CC methylation of DNA. Such in vitro methylation may be used to direct or
CC regulate DNA expression in biological systems, e.g. recombinant DNA
CC methylated in vitro may be introduced into a cell/organism to increase
CC or decrease expression of a desired polypeptide for which the native DNA
CC is under-methylated or not methylated. The polypeptides can also be
CC used to produce antibodies which are useful to detect and purify the
CC polypeptide or therapeutically e.g. to treat neoplastic disorders. The
CC polynucleotides are useful to produce probes and primers which are
CC useful diagnostically.
XX
SQ Sequence 4195 BP; 1061 A; 1059 C; 1123 G; 952 T; 0 other;
Query Match 18.1%; Score 757.8; DB 21; Length 4195;
Best Local Similarity 63.4%; Pred. No. 4.7e-174;
Matches 1250; Conservative 0; Mismatches 702; Indels 21; Gaps 5;
QY 991 GAGCCAGTAGAGGGGATGCTGGGGACAAGATGCTACCAAGCAGCCGACGATGAGCCT 1050
DB 875 GATCAGGAGGGTATGGATACCACAGGTGGATGAGAGCATATATGGAGACAGACA 934
QY 1051 GAGTATGAGGATGCGCGGGCTTTTGGCATTTGGAGAGCTGGTGGGGGAACTTCGGGGC 1110
DB 935 GAGTATCAGGATGATAAAGAGTTTGGATAGGTGACCTCGTGGGGGAAAGATCAAGGGC 994
QY 1111 TTCTCTGTGGCGCCAGCGCGAATTTGTGTTGGTGGATGACAGCGCGGAGCAGAGCT 1170
DB 995 TTCTCTGTGGCGCTGCCATGTTGGTGTCTCGAAGGCCACCTCCAGCGACAGGCCATG 1054
QY 1171 GAAGGCACTCGCTCGGTTCATGTTTCGAGATGCAAGTTCCTCAGTGTGTGTGGAG 1230
DB 1055 CCCGGAATGCGCTGGTACAGTGGTTTGGTGTGATGGCAAGTTTCTGAGATCTCTGCTGAC 1114
QY 1231 AAGCTCATGCGCTGAGCTCTCTTCGAGTGCATTTCCACAGGCCACCTACAACAGCAG 1290
DB 1115 AACTGTGCTCTGGGGCTGTTTCCAGCCAGCACTTTAATCTGGTACCTTCAATAGCTG 1174
QY 1291 CCCATGTACCGAAAGCCATCTACGAAGTCTCCAGTGGCCAGCAGCCGTGCGGGAAG 1350
DB 1175 GTTCTTTATAGGAAGGCCATGTACACACTCTGGAGAAAGCCAGGGTTTCAGTGGCAAG 1234
QY 1351 -CTGTTTCCAGCTTGGCATGACAGTGAAGTGAAG- - - - -TGGCAAGGCTGTGGAAG 1405
```

Db 1235 ACCCTTCACAGAGTCTCTGAGAGTCACTGAGAGACCAAGCTGAAGCCATGCTGAGTGG 1294  
 Oy 1406 TGCAGAACAGAGATGATTGAAATGGCCCTCGTGCTTCAGCCCTCGGGTCTTAAG 1465  
 Db 1295 GCCCAGCGGTGCTTCAAGCCTACTGAGGATCGAGGGCTCAAAACCAAGAAAGAACCA 1354  
 Oy 1466 GCGTGGAGCCACAGAAAGAAAGAAATCTTCAAGAAAGTTTACACCGACATGTGG 1525  
 Db 1355 GTGGTTAATAAGTCGAAGGTGCGTTCAGACAGTAACTTGAACCCAGAGAGCGC 1414  
 Oy 1526 TGGAGCTCGAAG-----CAGCTGCTTACGCCCCACCCCAACCCCAAGAAACC 1575  
 Db 1415 GAGAAACAAAGTCGAAGACGACAAACCAATGACTGCTGCTGAGAGTCCCCCACC 1474  
 Oy 1576 AGAAAGACCAACAGAAAGAAAGTCAAGGA---GATCATTTAGTACGCGACAAG 1632  
 Db 1475 AAGCGCTCAAGACAAATAGCTATGGCGGAAAGACCGAGGGAAGATGAGAGAGCGA 1534  
 Oy 1633 GAGGGCTGTGTATGAGGTGCGGCAAGAGTGCAGAAACATCGAGACATTTGTATCTCA 1692  
 Db 1535 GAACGGAATGGCTTCTCAAGTCAACAAACAAAGGCAATCGAAGACCGCTGTTGTCC 1594  
 Oy 1693 TGTGGGAGCTCAATGTCAACCTGAGACACCACTTTCATTTGAGGCGATGTGCCAAG 1752  
 Db 1595 TGTGGAAGAAAGAACCCCTGTGTCTTCCACCCCTCTTGAAGGTGGGCTGTGTCAAGT 1654  
 Oy 1753 TGTAAAGACTGCTTCTTGAAGTGTCTTACCAATGACGAGATGGGTACCACTCTAT 1812  
 Db 1655 TGCAGGATGCTTCTTGAAGTCTTCAATGATGATGAGAGCGGCTATGAGTCTTAC 1714  
 Oy 1813 TGCAACATCTGCTGTGGGGGGCGTGAAGTCTCATGTGTGGGAACAACTGCTGAG 1872  
 Db 1715 TGCAACGATGTGTGAGGGCGGTGMACTGTCTGTGACAGTAAACAAAGTGTCTGAGA 1774  
 Oy 1873 TGCTTTGTGTGAGTGTGTGATCTTGTGTGGGCGCAGAGTGTCTAGGAGCCATT 1932  
 Db 1775 TGCTTGTGTGAGTGTGTGAGTGTGTGTGGGCGCAGAGCAAGTGAAGTGTCCAG 1834  
 Oy 1933 AAGGAAGACCCCTGGAATCTCTAATGTGCGGCAATAGGSCACTTATGGCTGTCCGA 1992  
 Db 1835 CTGCAAGAACCTGTGAGCTCTAATATGTCTCTCCAGCGCTGCAATGGGGTCTCCGA 1894  
 Oy 1993 AGACGGGAAGACTGCGCTTCTGACTCCAGATGTCTTTGGCAATTAACCATGACC--AG 2049  
 Db 1895 CGAGGAAGATTTGAACATGCGCTGCAAGACTTCTTCACTAGATCTGTGAGAA 1954  
 Oy 2050 GAATTTGACCCCCCAAGGTTTACCCACTGTGCCAGCTGAGAAAGAAAGGACCCATCCGC 2109  
 Db 1955 GAAATTTGAGCCACCCAAAGTTGTACCCAGCAATTCCTGAGCCAAAGAGCCCATAGA 2014  
 Oy 2110 GTGCTGTCTCTTTGATGGATTTGTACAGGGCTCTGTGTGAAGAACCTGGGCATC 2169  
 Db 2015 GTCTGTCTCTGTGTGATGGAATTTGCAAGGGTACTTGTGTCAAGAGTGTGGTATT 2074  
 Oy 2170 CAAGTGAACCGTACATTTGCTCCGAGGTGTGAGAGTCTCCATCAACGTTGGCATGTG 2229  
 Db 2075 AAAGTGAAGAAAGTACATTTGCTCGAAGTCTGTGAGAGTCTCATGCTGTGGAACTGTT 2134  
 Oy 2230 CGGACACGGAAGATCATGTAGTGTGGGAGGTCCGAGCGTCAACAGAAAGCATATC 2289  
 Db 2135 AAGCATGAAGGCGAGATCAATATATGTCAATGACGTCCGAAATCAACAAAGAAATATT 2194  
 Oy 2290 CAGAGTGGGGCCCATTTGACCTGTGATTTGAGGCAAGTCCCTGCAATGACCTCCATT 2349  
 Db 2195 GAAAGATGGGGCCCGTTGACTTGTGTGTGTGAGAGCCCATGAAATGATCTCTTAAC 2254  
 Oy 2350 GTCAACCTGCGCGAAGGACTTATAGAGGTACTGCGCGCTCTTCTTGAAGTTCTAC 2409  
 Db 2255 GTCAATCTGCGCGAAAGGTTTATATAGAGGCAAGAAAGGCTTCTTCAAGTTTAC 2314  
 Oy 2410 CGCTCTGATGATGCGCGGCCAAGAGAGGAGATGATGCGCCCTTCTTGTGGCTCTT 2469  
 Db 2315 CACTTGTGATTTATACCGCGCCCAAGAGGCGGACAAACGCTTCTTGTGATGTTC 2374

Oy 2470 GAGATGTGTGGCCATGGCGCTTATGTACAAAGAGGACATCTTGCAATTTCTGATCT 2529  
 Db 2375 GAGATGTGTGGCCATGAAAGTGAATGACAAAGAAACATCTCAAGATTTCTTGCAATGT 2434  
 Oy 2530 AATCCCGTATGATTTGACCGCAAGAAAGTGTCTGTGACACAGAGGCCCGTTACTTCTGG 2589  
 Db 2435 AATCCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2494  
 Oy 2590 GGTATCTTCTGTCATGAAACAGCGCTTTGGCATCTGTGAATGATTAAGTGTGAGCTG 2649  
 Db 2495 GTTAACCTTACCCCGAATGAACAGCGCGTGTATGCTTCAAAAGATGATTAAGCTCGAGCTG 2554  
 Oy 2650 CAAGAGTGTGTGAGACGCGCAAGATAGCCAAAGTTGACAAAGTGTGAGACCAATTACCACC 2709  
 Db 2555 CAGGACTGTCTGAGAGTTCAGTAGGACAGCAAAAGTTAAAGAAAGTGCAGAAATTAACCAAC 2614  
 Oy 2710 AGGTCAACTCTATTAAGACAGGCGCAAAAGACCAAGCATTTCCCGCTTATGAAAGAGAG 2769  
 Db 2615 AAGTCGAATCTCATGACAGAGCGCAAAACCAAGCTTTTCCCTGTATGATGATGAGCAAG 2674  
 Oy 2770 GAGGACATCTGTGTGTGCTGATGAATGAAAGGTTTGGCTTCCCGTCCACTACACA 2829  
 Db 2675 GACGACGTTTGTGTGTGCTGATGAGCTGAAAGATCTTCCGCTTCCCTGCTCACTACACG 2734  
 Oy 2830 GACGCTTCCATGATGAGCGCGCTTGGCGAGCAGAGACTGTGGGCCATGCTGAGAGCGTG 2889  
 Db 2735 GACGCTTCCATGATGAGCG 2794  
 Oy 2890 CCGGTCAATCCCGACCTCTTGGCTCCCGCTGAAGAAATTTTGTGTGTGTA 2942  
 Db 2795 CCGGTCAATCAAGACACTGTTTGGCCCGCTTGAAGGACTACTTGTGCTGTGAATA 2847

RESULT 5  
 AAZ37098  
 ID AAZ37098 standard; DNA; 4145 BP.  
 XX  
 AC AAZ37098;  
 XX  
 DT 27-MAR-2000 (first entry)  
 XX  
 DB DNA encoding de novo DNA cytosine methyltransferase DNMT3B1.  
 XX  
 KW De novo DNA cytosine methyltransferase; DNMT3B1; neoplastic disorder;  
 KW carcinoma; sarcoma; leukaemia; DNA methylation; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key  
 FT CDS  
 FT  
 FT location/Qualifiers  
 FT 115..2676  
 FT /\*tag= a  
 FT /product= "de novo DNA cytosine methyltransferase  
 FT DNMT3B1"  
 FT  
 FT 1..108  
 FT /\*tag= b  
 FT /number= 1  
 FT 109..256  
 FT /\*tag= c  
 FT /number= 2  
 FT 257..318  
 FT /\*tag= d  
 FT /number= 3  
 FT 319..420  
 FT /\*tag= e  
 FT /number= 4  
 FT 421..546  
 FT /\*tag= f  
 FT /number= 5  
 FT 547..768  
 FT /\*tag= g  
 FT /number= 6  
 FT 769..927  
 FT exon



FT FT exon /tag= h  
FT FT exon /number= 7  
FT FT exon /tag= i  
FT FT exon /number= 8  
FT FT exon /tag= j  
FT FT exon /number= 9  
FT FT exon /tag= k  
FT FT exon /number= 10  
FT FT exon /tag= l  
FT FT exon /number= 11  
FT FT exon /tag= m  
FT FT exon /number= 12  
FT FT exon /tag= n  
FT FT exon /number= 13  
FT FT exon /tag= o  
FT FT exon /number= 14  
FT FT exon /tag= p  
FT FT exon /number= 15  
FT FT exon /tag= q  
FT FT exon /number= 16  
FT FT exon /tag= r  
FT FT exon /number= 17  
FT FT exon /tag= s  
FT FT exon /number= 18  
FT FT exon /tag= t  
FT FT exon /number= 19  
FT FT exon /tag= u  
FT FT exon /number= 20  
FT FT exon /tag= v  
FT FT exon /number= 21  
FT FT exon /tag= w  
FT FT exon /number= 22  
FT FT exon /tag= x  
FT FT exon /number= 23  
XX WO9967397-A1.  
XX 29-DEC-1999.  
XX 25-JUN-1999; 99WO-US14373.  
XX 25-JUN-1998; 98US-0090906.  
XX 24-JUL-1998; 98US-0093993.  
XX (GEO) GEN HOSPITAL CORP.  
XX Li E, Okano M, Xie S;  
XX WPI; 2000-106298/09.  
XX P-PSDB; AAY54058.  
XX New mouse and human polypeptides, useful to treat and diagnose  
XX neoplastic disorders e.g. carcinomas, sarcomas and leukemias -  
XX Example 1; Fig 1D; 114pp; English.  
XX The present sequence encodes a human de novo DNA cytosine

CC methyltransferase designated DNMT3B1. The DNMT3B gene also produces,  
CC through alternate splicing, at least two shorter isoforms termed  
CC DNMT3B2 (comprising amino acids 1-355 and 376-853 of AAY54058, see  
CC AAY54058, see AAY54060) (sequences not given in the specification).  
CC The polypeptides can be administered therapeutically, especially by  
CC expressing encoding polynucleotides, to treat diseases associated with  
CC DNA cytosine methyltransferase, such as neoplastic disorders e.g.  
CC carcinomas, sarcomas and leukemias. They can be used to diagnose, or  
CC determine susceptibility to neoplastic disorders, by assaying for  
CC polypeptide expression levels in mammalian cells/body fluids. They are  
CC useful to screen for compounds inhibiting/activating the polypeptide. The  
CC polypeptides can also be used for in vitro de novo methylation of DNA.  
CC Such in vitro methylation may be used to direct or regulate DNA  
CC expression in biological systems, e.g. recombinant DNA methylated in  
CC vitro may be introduced into a cell/organism to increase or decrease  
CC expression of a desired polypeptide for which the native DNA is  
CC under-methylated or not methylated. The polypeptides can also be used to  
CC produce antibodies which are useful to detect and purify the polypeptide  
CC or therapeutically e.g. to treat neoplastic disorders. The polypeptide  
CC polynucleotides are useful to produce probes and primers which are useful  
CC diagnostically.  
XX SQ Sequence 4145 BP; 1041 A; 1083 C; 1096 G; 925 T; 0 other;  
Query Match 17.9%; Score 749.8; DB 21; Length 4145;  
Best Local Similarity 62.5%; Pred. No. 4.2e-172;  
Matches 1234; Conservative 0; Mismatches 717; Indels 24; Gaps 3;  
QY 992 AGCCAGTAGAGGGGATGCTGGGGACAAGATGCTACCAAGACAGCCAGCTGAGCCTG 1051  
DB 701 AGCAGGGGGCATGGAGTCCCGCAGGTGGAGGAGACAGTGGAGATGAGAGTTCAG 760  
QY 1052 AGTATCAGGATGCGGGGGCTTTGGCATTTGAGAGCTGGTGTGGGGGAACTTCGGGGCT 1111  
DB 761 AGTATCAGGATGCGAAGGAGCTTTGGAATAGGGACCTCGTGTGGGGAAAGATCAAGGGCT 820  
QY 1112 TCTCCTGGTGGCCAGCCGAAATGTTGTCTTGGTGGATGACAGGCCGAGCCAGCAGCTG 1171  
DB 821 TCTCCTGGTGGCCCGCATGGTGGTGTCTTGGAAAGGCCACCTCCAAAGCGACAGGCTATGT 880  
QY 1172 AAGGCACCTCGCTGGGTCATGTGGTTCGGAGATGCAAGTCTCTCAGTGGTGTGTGGAGA 1231  
DB 881 CTGGCATGCGTGGTCCAGTGGTTTGGCGATGCAAGTCTCTCCGAGGTCTCTGCAGACA 940  
QY 1232 AGCTCATGCGCTGAGCTCCTCTGCGAGTGCATTCACCCAGCCACCTACACACAGCAGC 1291  
DB 941 AACTGGTGACCTGGGGCTGTTGAGCCAGCAGTCTTAATTTGGCCACCTTCAATAAGCTCG 1000  
QY 1292 CCATGTACCGCAAGGCCATCTACGAAGTCTCTCAGGTGGCCAGCAGCCGTGCCGGGAAG- 1350  
DB 1001 TCTCCTATCGAAAAGGCCATGTACCATGCTCTGGAGAAAGCTAGGTGCGAGCTGCAAGA 1060  
QY 1351 CTGTTTCCAGCTTGGCCATGACAGTGAAGTGAAGTGCAGTGGCAAGGC----- 1397  
DB 1061 CTTTCCAGCAGCCCTGGAGACTCATTTGGAGGACCACTGAAGCCCATGTTGGAGTGGG 1120  
QY 1398 ----TGTGGAAGTGCAGAACAGCAGATGATTGAATGGGCCCTCGGTGGCTTCCAGCCCT 1453  
DB 1121 CCACGGGGGCTTCAAGCCCACTGGGATCGAGGGCCCTCAAAACCCCAACACACGCAACCAG 1180  
QY 1454 CGGGTCTTAAGGGCTTGGAGCCACCAAGAAAGAGAAATCTTACAAGGAAGTTTACA 1513  
DB 1181 TGGTTAATAGTCGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1240  
QY 1514 CCGACATGGTGGGCTGAGCCTGAGCAGCTGTTAGCCGCCACCCACCAGCAGCAAGAAC 1573  
DB 1241 AGAACAGACTCGAAGACGACAGCAGCTGACGACTCAGCCACCTCTGACTACTCCCCGCAC 1300  
QY 1574 CCAGAAAGAGCACAACAGAGAAACCTAAGGTCAAGGAGATC-----ATTGATGAGCGCA 1627  
DB 1301 CCAAGCCCTCAAGACAAATTGCTATACACGCAAGACCGAGGGGATGAAGATCAGA 1360

QY 1628 CAAAGGAGCGCTGTGTATGAGGTGCGCCAGAGAGTGCAGAAATCGAGACATTGTA 1687  
 Db 1361 GCCGAGAACAAATGGCTTCAGATGTTGCCAACAGAGAGCTGGAGATGGCTGTT 1420  
 QY 1688 TCTCATGTGGAGACCTTCATGTCAACCTGAGACACCCACTCTTCAATTGAGGAGCATGGCC 1747  
 Db 1421 TGCTTGGGCGAGAAAAACCCGCTGCTTCCACCTCTCTTTGAGGGGGGCTCTGTC 1480  
 QY 1748 AGAAGTGAAGATCTCTTCTTGTGAGTGTCTTACCAAGTATGACAGCATGGGTACAGT 1807  
 Db 1481 AGACATGCCGGATGCTCTTCTTGTAGCTGTTTACATGTATGATGACATGGCTATCAGT 1540  
 QY 1808 CCTATTGACCATCTGCTGTGGGGGGCGTGAAGTGTCTCATGTGTGGAAACAACTGCT 1867  
 Db 1541 CTTACTGACTGTGTCTGTGGAGGGCCGAGAGCTGCTGCTTTGACGACAAACGAGCTGCT 1600  
 QY 1868 GCAAGTCTTTTGTGTGAGTGTGTGAATCTCTTGTGGGGCCAGAGCTGCTCAGGCG 1927  
 Db 1601 GCCGCTGTTTCTGTGTGAGTGTCTGAGAGTGTCTGTGGGCAACAGGCGAGCGCCGAGG 1660  
 QY 1928 CCAATTAAAGAAAGACCCCTGGAATCTGTACATGTCCGGGCATTAAGGGCACTTATGGCTGC 1987  
 Db 1661 CCAAGCTTCAGAGAGCCCTGAGAGCTGTACATGTGTCTCCGAGCGCTGTCTATGGCTCC 1720  
 QY 1988 TGGGAGACGAGGAGAGCTGGCTTCTCGACTCCAGATGTTCTTTGGCAATTAACATGACC 2047  
 Db 1721 TGGGCGCCGAGAGAGCTGGAACGTCGCTGACAGGCTTCTTCAACAGTACACGGGGC 1780  
 QY 2048 AGGAATTTGACCCCGCCAAAGTTTAAACCACTGTGCGAGCTGAGAAAGAAAGCCCATCC 2107  
 Db 1781 TTGAATACGAGAGCCCGCAACCTGTACCTGCACTTCCGAGCCCGAAGGCGGCCCATTC 1840  
 QY 2108 GCGTGTCTCTCTCTTTTGTATGAGATGCTTACAGAGGCTCTGTGTGAAAGAACTGGGCA 2167  
 Db 1841 GAGTCTGTCTATGTTTGTATGAGATGCTGAGAGCTGAGAGCTTCTTCTTCAAGAGTTGGCA 1900  
 QY 2168 TCCAGTGGACCGCTACATTTCCCTCGAGGTGTGAGAGACTCCATCAAGTGGGCAATGG 2227  
 Db 1901 TAAAGTGAAGAAAGTACGTGCTTCTGAAGTGTGAGAGTCCATTTGTTGGAACCG 1960  
 QY 2228 TGGCGCACAGGAGAAAGATCATGTACGTGGGAGCTCCGAGCGTCAACAGAAAGATA 2287  
 Db 1961 TGAAGCACAGGAGGAATATCAAAATACGTGAACGACGTGAGAGAACTCAAAAGAAATA 2020  
 QY 2288 TCCAGAGTGGGGCCCATTCGACCTGTGATTTGAGGAGCACTCCCTGCAATGACCTTCCA 2347  
 Db 2021 TTGAAGATGGGGCCCATTTGACTTGTGATTTGGCGAAGCCCATGCAAGATCTCTCAA 2080  
 QY 2348 TTGTCAACCTTGGCCGCAAGGACTTTATGAGGATCTGAGCGGCTCTTCTTGAATTC 2407  
 Db 2081 ATGTGAATCCAGCCAGGAAGGCTGTATGAGGGTACAGGCGGCTCTTCTTGAATTTT 2140  
 QY 2408 ACCGCTCTCTGATGATGCGCGGCCCAAGAGAGAGATGATGCCCTTCTTGTGCTCT 2467  
 Db 2141 ACCACCTGTGATTAATCTACGCGCCCAAGAGAGGATGACCGGCTCTTCTTGTGATGT 2200  
 QY 2468 TTGAAGATGTGTGCGCATGAGGCGCTTATGACAAAGAGGACATCTCGCATTTCTTGAGT 2527  
 Db 2201 TTGAAGATGTGTGAGCATTAAGGTGTGGGACAGAGGACATCTCAGGTTCTTGAGT 2260  
 QY 2528 CTAAACCCGCTGATGATGACGCAAGAGAGTCTGTGTGACACAGGCGCCGTTACTTCT 2587  
 Db 2261 GTATCCAGTGTATGATGATGACCAAGATTTCTGTGTGCTCAACAGGCGCCGATCTTCT 2320  
 QY 2588 GGGGTAACTTCTTGGCATTAACAGGCTTTGGCATCTCACTGTGATTAAGTGGAGC 2647  
 Db 2321 GGGGCAACCTTACCCGGGATGAACAGGCGCGTATAGCATCAAGATGATTAACCTGAGC 2380  
 QY 2648 TGCAAGATGTGTGAGGACAGGCGCAATTAAGCAAGTTCAAGAAAGTGAAGACATTACCA 2707  
 Db 2381 TGCAAGATGTGTGAGGACAGGCGCAATTAAGCAAGTTCAAGAAAGTGAAGACATTACCA 2440  
 QY 2708 CCAAGTCAAACTCTAATAAGAGGCGCAAGACGACATTTCCCGCTCTTCAATGAACGAGA 2767

Db 2441 CCAAGTGCAGTCTGATCAACAGAGGAGAAAAACCACTTTCCCTGTTCTCATGAATGCA 2500  
 QY 2768 AGGAGACATCTCTGTGTGACATGAAATGAAAGGTTTGGCTTCCCGTCCACTACA 2827  
 Db 2501 AAGAGATGTTTGTGTGCTGACGTGAGAGATCTTTGGCTTCTGTGCACTTACA 2560  
 QY 2828 CAGAGCTCTCCAAATGAGCCGCTTGGCGAGAGCAGAGACTGCTGGCCGATGTGAGCG 2887  
 Db 2561 CAGAGCTCTCCAAATGAGCCGCTTGGCGAGAGCAGAGACTGCTGGGAAAGTCTGAGCG 2620  
 QY 2888 TGGCGTCAATCCGACCTTCTGCTCCGCTGAAAGAAATATTTGCTTGTGTA 2942  
 Db 2621 TGCTGTATCTCGACACCTTCTGCCCCCTTGAAGAGACTTGTGATGATA 2675  
 RESULT 6  
 AB057540  
 ID AB057540 standard; cDNA; 709 BP.  
 AC AB057540;  
 DT 02-AUG-2002 (first entry)  
 DE Human colon cancer related nucleotide sequence SEQ ID NO:1235.  
 DE Human; colon cancer; cancer; tissue profiling; forensic; mapping;  
 KW genetic analysis; diagnostic; antisense therapy; gene; ss.  
 OS Homo sapiens.  
 PN WO200229086-A2.  
 PD 11-APR-2002.  
 XX 02-OCT-2001; 2001WO-US30732.  
 XX 02-OCT-2001; 2000US-237271P.  
 PR (FARB ) BAYER CORP.  
 PA Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;  
 PI Thiglingam A, Lewis ME;  
 XX WPI; 2002-426115/45.  
 DR New isolated nucleic acid that is differentially expressed in cancer  
 PT tissues useful for determining the presence of colon cancer in a cell  
 PT or tissue type, and in antisense therapy  
 XX Claim 1; Fig 1; 796pp; English.  
 XX AB056306 to AB060787 represent isolated nucleic acids (I) differentially  
 CC expressed in cancer tissues. ABR78993 to ABR79004 represent proteins  
 CC encoded by the AB060776 to AB060787 nucleic acid sequences. (I) can be  
 CC used in antisense therapy. An antibody immunoreactive with a polypeptide  
 CC encoded by (I) is useful for detecting cancer in a patient sample, and  
 CC for detecting the presence or absence of a polynucleotide encoded by a  
 CC nucleic acid which hybridizes to (I) in a cell. A probe/primer derived  
 CC from (I) can be used for determining the presence of a nucleic acid which  
 CC hybridizes to (I), and for determining the phenotype of cells in a sample  
 CC of cells from a patient. (I) is useful for determining the presence of  
 CC colon cancer in a cell or tissue type, for determining the presence or  
 CC state of other type of cancer, in antisense therapy, to generate  
 CC macroarrays on a solid surface, to identify a chromosome on which the  
 CC corresponding gene resides, and in tissue profiling, forensics, genetic  
 CC analysis, mapping and diagnostic applications. (I) can be used to raise  
 CC antibodies, and to screen for peptide analogues and antagonists.  
 XX  
 XX Sequence 709 BP; 163 A; 177 C; 193 G; 168 T; 8 other;  
 Query Match 13.3%; Score 559; DB 24; Length 709;  
 Best Local Similarity 88.8%; Pred. No. 5.5e-126;

Matches 624; Conservative 0; Mismatches 73; Indels 6; Gaps 2;									
QY	2250	GTACGTCGGGACGTCGCGAGGTCACACAGAGCATATCCAGAGTGGGGCCCATTCGA	2309						
Db	1	GGACGTCGGGACGTCGCGAGGTCACACAGAGCATATCCAGAGTGGGGCCCATTCGA	60						
QY	2310	CCTGTGATTTGGAGGAGTCCCTGCAATGACCTCTCCATTGTCAACCTCCCGCAAGGG	2369						
Db	61	TCTGTGATTTGGGGGAGTCCCTGCAATGACCTCTCCATTGTCAACCTCCCGCAAGGG	120						
QY	2370	ACTTTATGAGGTCATGGCGGCTCTCTTTGAGTTCTACCGCTCTCTGATGATGGCG	2429						
Db	121	CCTCTACAGGGCACTGGCGGCTCTCTTTGAGTTCTACCGCTCTCTGATGATGGCG	180						
QY	2430	GGCCAGGAGGAGATGATGGCCCTCTCTCTGCTCTTTGAGATGTGGTGGCCATGGG	2489						
Db	181	GGCCAGGAGGAGATGATGGCCCTCTCTCTGCTCTTTGAGATGTGGTGGCCATGGG	240						
QY	2490	CCTTAGTGACAGAGGAGATCTCGGATTTCTTGAGTCTTAACCCCTGATGATGAGC	2549						
Db	241	CCTTAGTGACAGAGGAGATCTCGGATTTCTTGAGTCTTAACCCCTGATGATGAGC	300						
QY	2550	CAAGAAGTGTCTGTCACACAGGGCCCTTACTTCTGGGTTAACTTCTCTGGCATGA	2609						
Db	301	CAAGAAGTGTCTGTCACACAGGGCCCTTACTTCTGGGTTAACTTCTCTGGCATGA	360						
QY	2610	CAGGCTTTTGGCATCCACTGTGAATGATAAGCTGGAGCTGCAAGAGTGTCTGGAGCAGG	2669						
Db	361	CAGGCTTTTGGCATCCACTGTGAATGATAAGCTGGAGCTGCAAGAGTGTCTGGAGCAGG	420						
QY	2670	CAGATAGCCAGTTCAGCAAGTGGAGCATTACCACTGAGTCAAACTCTATAAGCA	2729						
Db	421	CAGATAGCCAGTTCAGCAAGTGGAGCATTACCACTGAGTCAAACTCTATAAGCA	480						
QY	2730	GGCAAGAGCAGCATTTCCCGTCTTCATGAACGAGAGGAGCATCTGTGTGT-GCA	2788						
Db	481	GGCAAGAGCAGCATTTCCCGTCTTCATGAATGAAGAGAGGAGCATCTGTGTGTGCA	540						
QY	2789	CTGAATGGAAGGTTGTTGGCTTCCCGTCCACTACACAGAGCTCTCCACATGAGCC	2848						
Db	541	CTGAATGGAAGGTTGTTGGTTCCCGTCCACTACACAGAGCTCTCCACATGAGCC	600						
QY	2849	GCTTGGCGAGGAGAGTGTGGCGGATCGT-----GGAGCGTGGCGGTATCGGCA	2903						
Db	601	GCTTGGCGAGGAGAGTGTGGCGGATCGT-----GGAGCGTGGCGGTATCGGCA	660						
QY	2904	CCTTCTCGCTCCCGTGAAGGAATATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	2946						
Db	661	TNTTTNGNTTCCGTTGAAGGAATATTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	703						
RESULT 7									
AAH33860									
ID	AAH33860	standard; cDNA; 1060 BP.							
XX	AAH33860;								
AC	AAH33860;								
XX	03-SEP-2001 (first entry)								
DT	03-SEP-2001 (first entry)								
XX	Human colon cancer antigen encoding cDNA SEQ ID NO:916.								
DE	Human colon cancer antigen encoding cDNA SEQ ID NO:916.								
XX	Human; colon cancer; colon cancer antigen; diagnosis; detection;								
KW	Human; colon cancer; colon cancer antigen; diagnosis; detection;								
XX	colorectal carcinoma; ss.								
OS	Homo sapiens.								
XX	Human								
XX	Human								
PN	WO200122920-A2.								
XX	05-APR-2001.								
PD	28-SEP-2000; 2000WO-US26524.								
XX	28-SEP-2000; 2000WO-US26524.								
PF	28-SEP-2000; 2000WO-US26524.								
XX	28-SEP-2000; 2000WO-US26524.								
XX	29-SEP-1999; 99US-0157137.								
PR	29-SEP-1999; 99US-0157137.								

PR	03-NOV-1999; 99US-0163280.								
XX	(HUMA-) HUMAN GENOME SCI INC.								
DA	Ruben SM, Barash SC, Birse CE, Rosen CA;								
XX	WPI; 2001-235357/24.								
DR	P-PSDB; AAG74429.								
XX	Nucleic acids encoding 4277 human colon cancer-associated polypeptides,								
PT	useful for preventing, diagnosing and/or treating colorectal cancers -								
XX	Claim 1; Page 2829; 9803pp; English.								
XX	AAH32943 to AAH37195 and AAG77788 represent human colon								
CC	cancer-associated nucleic acid molecules (N) and proteins (P), where								
CC	the proteins are collectively known as colon cancer antigens. The colon								
CC	cancer antigens have cytostatic activity and can be used in gene								
CC	therapy and vaccine production. N and P may be used in the prevention,								
CC	diagnosis and treatment of diseases associated with inappropiate P								
CC	expression. For example, N and P may be used to treat disorders								
CC	associated with decreased expression by rectifying mutations or deletions								
CC	in a patient's genome that affect the activity of P by expressing								
CC	inactive proteins or to supplement the patient's own production of P.								
CC	Additionally, N may be used to produce the colon cancer-associated ps,								
CC	by inserting the nucleic acids into a host cell and culturing the cell								
CC	to express the proteins N and P can be used in the prevention, diagnosis								
CC	and treatment of colorectal carcinomas and cancers. AAH37195 to AAH37204								
CC	present invention.								
CC	N.B. Pages 666 to 682 and page 7053 of the sequence listing were								
CC	missing at time of publication, meaning no sequences are present for								
CC	SEQ ID NO:1027 to 1052, 7921 and 7922.								
XX	Sequence 1060 BP; 267 A; 262 C; 297 G; 227 T; 7 other;								
QY	Query Match	13.1%;	Score 547.6;	DB 22;	Length 1060;				
Db	Best Local Similarity	72.3%;	Pred. No. 4.1e-123;						
QY	Matches 703; Conservative	4;	Mismatches 266;	Indels	0;	Gaps	0;		
QY	1970	AGGGACCTATGCGCTGCTCGAAGACGGGAAGACTGGCCCTTCTCGACTCCACAGATGTTCT	2029						
Db	9	AGCGTGTCTATGCGCTGCTCGGCGCCGGAAGGACTGGAACGTGCGCTGCGAGCCTCTCT	68						
QY	2030	TTCCCAATACCATGACACAGCAATTTGACCCCAAGGTTTACCCAGCTGTGCCAGCTG	2089						
Db	69	TCACCATGACACGGGCTTGAATACGAAGCCCAAGCTGTACCTGCAATTCGCCGAG	128						
QY	2090	AGAGAGGAGGAGCCATCGCGTGTCTCTCTTTGATGGGATGCTACAGGCTCTCTGG	2149						
Db	129	CCGAGGCGGCCCTTTCGAGTCTCTGTCATTTGATGGCATCGGACAGCTACCTAG	188						
QY	2150	TGCTGAAGGACCTGGGATCCAGTGGACCGCTACATTTGCTCCGAGGTGTGTGAGACT	2209						
Db	189	TCCTCAAGAGTGGGATTAAGGTAGGAAAGTACGTGCTTCTGAAGTGTGTGAGAGT	248						
QY	2210	CCATCACGTTGGGATGTCGGGACCAAGGAAAGATCATGTACGTGGGGAGCTCCGCA	2269						
Db	249	CCATTGCTGTGGAAACCGTGAAGCAGAGGGGAAATATCAATACGTGAACGATGTGAGGA	308						
QY	2270	GCCTCACACAGAGCATATCCAGGAGTGGGCGCCATTGACCTGTGTGATTTGGAGGAGTC	2329						
Db	309	ACATCACAAAGAAATATTTGAAGAAATGGGCGCCCTTGTGATTTGGCGGAAGCC	368						
QY	2330	CCTGCAATGACCTCTCCATTTGTCAACCTGTCGCGCAAGGAGCTTTATGAGGTTACTGGCC	2389						
Db	369	CATGCAAGCATCTCTCAATGTGAATCCAGCCAGGAAAGGCTGTATGAGGGTACAGGCC	428						
QY	2390	GCCTCTCTTTGAGTCTTACCGCTCTCTGATGTCGCGGCGCCCAAGGAGGAGATGATC	2449						
Db	429	GGCTCTCTTTGAGTCTTACCGCTCTCTGATGTCGCGGCGCCCAAGGAGGAGGATGATC	488						
QY	2450	GGCTCTCTTTGAGTCTTACCGCTCTCTGATGTCGCGGCGCCCAAGGAGGAGGATGATC	2509						

Db	489	GGCGCTTCTTCTGATGATTGTGACATGTTGMASSCATGAAAGTTGGCGACAAAGAGGACA	548
QY	2510	TCTGCGATTTCTTGTAGTCTPAACCCCGTGTATGTGACGCCAAAGAGTGTCTGTGCAC	2565
Db	549	TCTCAACGGTTCCTGGAGTGTATCCASTGTATGTATGCCATCAAAAGTTTCTGTCTGC	608
QY	2570	ACAGAGCCCGTAACTTCTTGGGGTAACTTCTCGCATGAACAGCCCTTGGCATCCACTG	2629
Db	609	ACAGGGCCCGATACTTCTGGGGCAACTTACCCGGATGAACAGGCCCGTGTATGATCA	668
QY	2630	TGAATGATTAAGCTGGAGCTGCAAGAGTGTCTGGAGCAGGCGAATATGCAAGTTTCAGCA	2689
Db	669	AGAAATGATTAACCTCGNGCTGCAGAGACTGTCTGGAATACAAATAGGATATGCAAGTTAAGA	728
QY	2690	AAGTAGAGACATTAACCAACAGSTCAACTTAATTAAGCAGGCGAAAGACCAAGCATTTCC	2749
QY	729	AAGTACAGACAAATPAACCAACCAAGTCGAATCGATCAAAACAGGGGAAAAAACCAACTTTTCC	788
QY	2750	CCGTCTTCATGAACGAGAAGGAGACATCCTGTGGTGCACCTGAAATGSAAGGGTGTGG	2809
Db	789	CTGTGTGATGAATGGCAAGAGAGATGTTTNGTGTGCACCTGAGCTGMAAGNFTCTTGG	848
QY	2810	GCTTCCCGGTCCACTACACAGACGTCTTCAACATBAGCCGCTGGCCAGGCGAGACTGC	2865
Db	849	GCTTTCCTGTGACTPAACAGACGTCGTCCAACATGGGCCGTGTGGCCCGCCAGAACTGC	908
QY	2870	TGGGCGATCGTGGAGCGTGCAGGTGATCGGCCACTCTTGGCTCCGCTGAAGGAATATT	2929
Db	909	TGGGAAGGTCCTGGAGCGTGCCTGTATCGACACACTCTTGGCCCTCTTGAAGGACTACT	968
QY	2930	TTGCTTGTGTGA 2942	
Db	969	TTGCATGTGATA 981	

RESULT	8
AAH14009	
ID	AAH14009 standard; cDNA; 3017 BP.
XX	
AC	AAH14009;
XX	
DT	26-JUN-2001 (first entry)
XX	
DE	Human cDNA sequence SEQ ID NO:11099.

Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS	Homo sapiens.
XX	
XX	EP1074617-A2.
PN	
XX	
PD	07-FEB-2001.
XX	
XX	
PF	28-JUL-2000; 2000EP-0116126.
XX	
XX	
PR	29-JUL-1999; 99JP-0248036.
PR	27-AUG-1999; 99JP-030253.
PR	11-JAN-2000; 2000JP-018776.
PR	02-MAY-2000; 2000JP-0183767.
PR	09-JUN-2000; 2000JP-0241899.
XX	
PA	(HELI-) HELIX RES INST.
XX	
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saïto K, Yamamoto J,
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX	
DR	WPI; 2001-318749/34.

PT Primers for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -

XX Claim 8; SEQ ID 11099; 2537bp + CD ROM; English.  
PS  
XX  
CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB93893 represent human amino acid sequences; and AAH13829 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
SQ Sequence 3017 BP; 749 A; 751 C; 760 G; 757 T; 0 other;

Query Match	11.5%	Score 482.8;	DB 22;	Length 3017;
Best Local Similarity	67.08;	Pred. No. 4.5e-107;		
Matches 685;	Conservative	0;	Mismatches 337;	Indels 0;
				Gaps 0;

QY	1617	TGATGAGCCACAAAGGAGCGCGTGGTGTATGAGGTGGCCAGMAGTGACAAAATCGA	1616
Db	429	TGAAGATCAGAGCCGAGAACAAATGGCTTAGATGTTCACCAACAAGACAGCCTTGA	488
QY	1677	GGACATTGTATCTCATGTGGGAGGCTCAATGTATCCCTGAGSAGCACCACCTTCAATGG	1736
Db	489	AGATGCGCTTGTCTTGTGGCAGGAAAAACCCTGTGTCCTTCCACCCTCTTGTAGGG	548
QY	1737	AGGCATGTGCCAAGACTGTAAAGAACTGTTCTTGGAGTGTGCTTACCAGTATGACAGCA	1796
Db	549	GGGGCTGTGTACAGACATGCCGGGATGCGCTTCTTGAGCTGTTTTACTGTATGATGACGA	608
QY	1797	TGGGTACCAGTCTTATTTGACACATCTGCTGTGGGGGCGGTGAAGTGTCAATGTGGAA	1856
Db	609	TGGCATACGTCTTACTGTGCACTGTGTGTGTCGAGGGCGGAGAGCTGCTGTTCAGCAA	668
QY	1857	CAACAACCTGCTGAGGTGCTTTTGTGTGAGTGTGTGATCTTGTGTGGGGCCAGAGC	1916
Db	669	CACGAGCTGCTCGCGGTGTTTCTGTGTGAGTGTGCTGAGGTGTGCTGGTGGGCACAGGCAC	728
QY	1917	TGCTCAGCGACCATTAAGGAAGACCCTTGGAATCTGTAATGTGCGGGCATTAAGGGCAC	1976
Db	729	AGCGGCCGAGGCCAAGCTTACAGAGCCCTGAGAGCTGTACAATGTGTCTCCGACAGCTG	788
QY	1977	CTATGGCTGCTGTGCGAAGACGGGAAGACTGGCCTTCTTGACTTCAGATGTTCTTTGGCCA	2036
Db	789	TCAATGGCGTCTGTGCGGCGCCGAGAGACTGGAAAGTGTGCGCTTGACAGCCTTCTTCAACAG	848
QY	2037	TAACCATGACAGGAATTTGACCCCCCAAAAGTTTACCACCTGTGCAGCTTGACAAGAG	2096
Db	849	TGGCAGCGGGCTTTGAATACGAAGCCCCCAAGCTGTACCTGTGCATACCCGAGCCCGAAG	908
QY	2097	GAAGCCCATTCGCGGTGCTCTCTCTTGAATGGATTGTCTACAGGGCTCCTGGTGTGAA	2156
Db	909	GCGGCCCATTCGAGTCTCTGTATTTGTGTGATGTGCATCGGACAGGGCTACCTAATGCTCTCA	968
QY	2157	GGACCTGGGCATCAAGTGAACCGCTACATTGCTCCGAGTGTGTGAGGACTCATCAC	2216
Db	969	AGAGTTGGGCATTAAGGTAGAAAAGTACGTGCTTTGAAAGTGTGTGAGGAGTCCATTGC	1028
QY	2217	GGTGGGCATGTGTGCGGACCAAGGAAAGATCATGTATACGTGGGGGACGTCCGACAGCTAC	2276

Db	1029	TGTTGGAACCGTGAAGCAGAGGGGAATATCAATACGTGAACGACGTGAGGAACATCAC	1088	PR	14-AUG-2000;	2000US-0225213.
Qy	2277	ACAGAAAGCATATCCAGAGTGGGGCCCAATCGACCTGGTGAATGGAGGACGTCCCTGCAA	2336	PR	14-AUG-2000;	2000US-0225214.
Db	1089	AAAGAAATATGAAGATGGGCCCATTTGACTTGGTGAATGGCGGAAGCCCATGCAA	1148	PR	14-AUG-2000;	2000US-0225267.
Qy	2337	TGACCTCTCCATGTCAACCTCCCGCAAGGACCTTTATGAGGGTACTGGCCGCTCTT	2396	PR	14-AUG-2000;	2000US-0225268.
Db	1149	CGATCTCTCAATGTGAATCCAGCCAGGAAGGCTGTATGAGGGTACAGCCCGGCTCTT	1208	PR	14-AUG-2000;	2000US-0225270.
Qy	2397	CTTTGAGTTCTACCGGCTCTCTGCATGATGCGCGGCCCAAGAGGAGATGATCGCCCTT	2456	PR	14-AUG-2000;	2000US-0225447.
Db	1209	CTTCGAATTTTACCACCTGCTGAATTAATCAAGCCCAAGAGGGTGAATGACCGGCCCTT	1268	PR	14-AUG-2000;	2000US-0225757.
Qy	2457	CTTCTGGCTCTTTGAGAAATGTGTGGCCATGGGGCTTGTAGTACAGAGGGACATCTCGG	2516	PR	14-AUG-2000;	2000US-0225758.
Db	1269	CTTCTGGATGTTTGAAGATGTTGTAGCCATGAAGTTGGCCAGAGGGACATCTCAG	1328	PR	14-AUG-2000;	2000US-0225759.
Qy	2517	ATTTCTTGTAGTTAAACCCGCTGATGATGACGCCAAGAAAGTGTCTGTGCACAGGGC	2576	PR	18-AUG-2000;	2000US-0226279.
Db	1329	GTTCTGGAGTGAATCCAGTGAATGATGATGATGCCATCAAAAGTTTCTGTGTCTCAGGGC	1388	PR	22-AUG-2000;	2000US-0226681.
Qy	2577	CGTTTACTTCTGGGTAACTTCTCTGCGATGAACAGGCTTTGGCATCCACTGTGAATGA	2636	PR	22-AUG-2000;	2000US-0226868.
Db	1389	CGATACTTCTGGGCAACCTACCGGGATGAACAGGATCTTTGGCTTCTCTGTGCACTA	1448	PR	23-AUG-2000;	2000US-0227009.
Qy	2637	TA 2638		PR	30-AUG-2000;	2000US-0228924.
Db	1449	CA 1450		PR	01-SEP-2000;	2000US-0229287.
RESULT 9						
ID	AAL00271 standard; cdna; 622 BP.					
AC	AAL00271;					
DT	21-NOV-2001 (first entry)					
DE	Human reproductive system related antigen cdna SEQ ID NO: 272.					
KW	Human; reproductive system related antigen; reproductive system disorder;					
KW	cancer; gene therapy; ss.					
OS	Homo sapiens.					
PN	WO200155320-A2.					
PD	02-AUG-2001.					
PF	17-JAN-2001; 2001WO-US01339.					
PR	31-JAN-2000; 2000US-0179065.					
PR	04-FEB-2000; 2000US-0180628.					
PR	24-FEB-2000; 2000US-0184664.					
PR	02-MAR-2000; 2000US-0186350.					
PR	16-MAR-2000; 2000US-0189874.					
PR	17-MAR-2000; 2000US-0190076.					
PR	18-APR-2000; 2000US-0198123.					
PR	19-MAY-2000; 2000US-0205515.					
PR	07-JUN-2000; 2000US-0209467.					
PR	28-JUN-2000; 2000US-0214886.					
PR	30-JUN-2000; 2000US-0215135.					
PR	07-JUL-2000; 2000US-0216647.					
PR	07-JUL-2000; 2000US-0216880.					
PR	11-JUL-2000; 2000US-0217487.					
PR	11-JUL-2000; 2000US-0217496.					
PR	26-JUL-2000; 2000US-0218290.					
PR	26-JUL-2000; 2000US-0220963.					
PR	26-JUL-2000; 2000US-0220964.					
PR	14-AUG-2000; 2000US-0224518.					
PR	14-AUG-2000; 2000US-0224519.					

PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246509.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251889.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
Rosen CA, Barash SC, Ruben SM;  
WPI; 2001-465570/50.  
DR P-PSDB; AAM94301.  
XX  
XX  
PT Isolated nucleic acid molecule encoding a reproductive system antigen  
PT is used in preventing, treating or ameliorating a medical condition -  
XX  
XX  
XX Claim 1; SEQ ID NO 272; 1297bp + Sequence Listing; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
CC number of human reproductive system related antigens. These can be used  
CC in the prevention and treatment of reproductive system disorders,  
CC including cancer. The present sequence is a coding sequence of the  
CC invention.  
XX  
XX  
SQ Sequence 622 BP; 154 A; 161 C; 189 G; 112 T; 6 other;  
Query Match 9.9%; Score 413; DB 22; Length 622;  
Best Local Similarity 87.7%; Pred. No. 2.1e-90;  
Matches 471; Conservative 2; Mismatches 62; Indels 2; Gaps 2;  
OY 1446 CCAAGCCCTCGGCTCTTAAGGCGCTGAGCCACCAAGAAAGAAATCTTACAAGGA 1505  
Db 2 CCAAGCCTTCTGCGCCCTTAAGGCGCTTAAGCCACCAAGAAAGAAATCTTACAAGGA 61  
OY 1506 AGTTTACACCGACATGTGGGTGGAGGCTGAAGCAGCTGCTTACGCCCCACCCACCAGC 1565

Db 62 AGTGTACCGGACATGTGGGTGGAGCTTGAGCAGCTGCTTACGACCACTTCCACAGC 121  
OY 1566 CAAGAAACCCAGAAAGAGCACAACAGAGAAACCTTAAGGTCAAGAGATCATGTAGAGC 1625  
Db 122 CAAGAAAGCCCCGAAAGAGCACAAGCGAGAGAGCCCAAGGTCAAGAGATTAATGATGAGC 181  
OY 1626 CACAAGGAGCGGCTGTGTATGAGGTGCGCCAGAAAGTGCAAGAAATCGAGAGCATTTG 1685  
Db 182 CACAAGAGAGCGGCTGTGTATGAGGTGCGCCAGAAAGTGCGGAACATTGAGAGCATCTG 241  
OY 1686 TATCTCATGTGGAGAGCCTCAATGTCAACCTTGAGAGCACCCTTTCATTGAGAGCATGTG 1745  
Db 242 CATCTCTGTGGAGAGCTCAATGTATTAACCTTGAAACACCCCTTCTTGTGAGGAATGTG 301  
OY 1746 CCAAGACTGTAAAGACTGCTTCTTGAGAGTGTCTTACCAATATGACGAGATGGTACCA 1805  
Db 302 CCAAAACTGCAAGAACTGCTTCTTGAGAGTGTCTTACCAATATGACGAGATGGTACCA 361  
OY 1806 GTCTTATTCACACCATCTGTGTGGGGGCGTGAAGTGTCTCATGTGTGGAAACAACACTG 1865  
Db 362 GTCTTACTGCACCATCTGTGTGGGGGCGTGAAGTGTCTCATGTGTGGAAACAACACTG 421  
OY 1866 CTGCAGGTGCTTTTGTGTGAGTGTGTGATCTCTTGTGGGGCGAGAGCTGTCAAGC 1925  
Db 422 CTGCAGGTGCTTTTGTGTGAGTGTGTGATCTCTTGTGGGGCGAGAGCTGTCAAGC 480  
OY 1926 AGCCATTAAAGAAAGCCCTGGAACCTGCTACATGTGCGGCAATTAAGGCACTATG 1982  
Db 481 AGSCATTAAAGAAAGCCCTGGAACCTGCTACATGTGCGGCAATTAAGGCACTATG 536

## RESULT 10

AAS86165

ID AAS86165 standard; cDNA; 711 BP.

XX AAS86165;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #21969.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW Food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX P-PSDB; ABG21978.

XX New isolated polynucleotide and encoded polypeptides, useful in

XX diagnostics, forensics, gene mapping, identification of mutations

XX responsible for genetic disorders or other traits and to assess

XX Claim 1; SEQ ID NO 21969; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and

XX polypeptide (II) sequences. (I) is useful as hybridisation probes,

XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 711 BP; 181 A; 164 C; 199 G; 167 T; 0 other;

Query Match 8.1%; Score 341; DB 23; Length 711;  
Best Local Similarity 74.7%; Pred. No. 7.9e-73;  
Matches 428; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 2110 GTGCTGCTCTCTTTGATGGATTGCTACAGGGCTCTGCTGTGTAAGGACCTGGGCATC 2169

Db 1 GTCTGCTCATTTGTTGATGGCATCGCAGAGGTACTAGTCTCAAGAGTTGGGCATA 60

QY 2170 CAAGTGGACCGTACATTCCTCCGAGGTGTGTGAGGACTCCATCAGGTGGGATGGTG 2229

Db 61 AAGTGTAGAAAGTACGCTCTTGAAGTGTGTGAGGAGTCCATTTGTTGGAACCGTG 120

QY 2230 CGGCACAGGGAAGATCATGTACGTCCGGGACGTCCGACGGTCAACAGAGCATATC 2289

Db 121 AAGCAGAGGGGATATCAATACGTCAACGCTGAGGAAACATCAACAGAAAATATT 180

QY 2290 CAGGAGTGGGGCCCATTCGACCTGTGTGATGGAGGCAATCCCTCCATGACCTCTCCATT 2349

Db 181 GAAGAATGGGGCCCATTTGACTTGGTGTATGGCGGAAGCCCATGCAACGATCTCTCAAAAT 240

QY 2350 GTCAACCTCTGCCCGAAGGACATTTATGAGGCTACTGCGCGCTCTTCTTTGAGTTCTAC 2409

Db 241 GTGAATCAGCCAGCAAGAAAGCCCTGTATAGGGTACAGCGCGCTCTTCTCGAATTTTAC 300

QY 2410 CGCTCTGTCATGATCGGGCCCAAGAGGAGATGATCGCCCTTCTTCTGCTCTTT 2469

Db 301 CACCTGCTGAATTTACTCACGCCCCCAAGAGGGGTGATGACCGCGCTCTTCTTGATGTTT 360

QY 2470 GAGAAATGGTGCCCATGGCGGTAGTGACAAGAGGACATCTCGGATTTCTTGAGTCT 2529

Db 361 GAGAAATGTTGTAGCCATGAAGTTGGCGACAAGAGGAGACATCTCACGGTTCTCTGAGTGT 420

QY 2530 AACCCGCTGATGATGACCGCAAGAGTCTCTGCTCACACAGGGCCCGTTACTTCTGG 2589

Db 421 AATCCAGTATGATGATGCCATCAAAAGTTTCTGCTCTCACAGGGCCCGATCTTCTGG 480

QY 2590 GGTAACTCTCTGGCATGAACAGGCTTTGGCATTCCTACTGTGAATGATAAGCTGAGCTG 2649

Db 481 GGCACCTACCCGGGATGAACAGGCCCGTGTATAGCATCAAGAATGATAAATCTGAGCTG 540

QY 2650 CAAGAGTGTCTGAGCACCGCAGAAATAGCCAAG 2682

Db 541 CAGGACTGCTTGGATACAAATAGGATAGCCAAG 573

RESULT 11

AAF80537 standard; cDNA; 2077 BP.

XX

AC AAF80537;

XX 08-JUN-2001 (first entry)

XX Receptor #25 partial coding sequence.  
DE  
XX Probe; microarray; cancer; immunopathology; neuropathology; ss.  
XX Balaena mysticetus.  
XX US6183968-B1.  
XX 06-FEB-2001.  
XX 25-MAR-1999; 99US-0276531.  
XX 27-MAR-1998; 98US-0079677.  
XX (INCY-) INCYTE PHARM INC.  
XX Bandman O, Lal P, Hillman JL, Yue H, Reddy R, Guegler KJ;  
XX Baughn MR;  
XX WPI; 2001-201999/20.  
XX Composition having probes which comprise part of gene sequence encoding  
XX proteins associated with cell proliferation useful as hybridizable  
XX array elements in Microarrays to monitor expression of target  
XX polynucleotide  
XX Claim 1; Columns 93-96; 104pp; English.  
XX The present invention relates to a composition comprising several  
XX polynucleotide probes. Probes can be derived from the present sequence.  
XX The probes are immobilised and are preferably useful as hybridisable  
XX array elements in a microarray for monitoring the expression of several  
XX polynucleotides. The microarray can be used in the diagnosis of cancers  
XX such as adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma, sarcoma  
XX and tetracarcinoma etc., immunopathology such as AIDS, Addison's  
XX disease, adult respiratory distress syndrome, allergies, anaemia, asthma,  
XX atherosclerosis and bronchitis etc., neuropathology such as Alzheimer's  
XX disease, amnesia, amyotrophic lateral sclerosis, bipolar disorder,  
XX catatonia and cerebral neoplasms etc. The microarray can also be used to  
XX investigate an individual's predisposition to a disease such as cancer,  
XX immunopathology or neuropathology. Also, the microarray can be used for  
XX investigating cellular response to infection, drug treatment etc. The  
XX microarray can be used for diagnostics, prognostics and treatment  
XX regimens, drug discovery and development, toxicological and  
XX carcinogenicity studies, forensics, pharmacogenomics etc. The array can  
XX also be used for monitoring disease progression.

XX Sequence 2077 BP; 539 A; 485 C; 477 G; 567 T; 9 other;

Query Match 7.7%; Score 324.6; DB 22; Length 2077;  
Best Local Similarity 74.8%; Pred. No. 1.4e-68;  
Matches 416; Conservative 0; Mismatches 139; Indels 1; Gaps 1;

QY 2387 GCGCCTCTCTTTGATGTTCTACCGCTCTCTGTCATGATCGCGGCCCAAGGAGGATG 2446

Db 62 GCGCGCTCTCTTCTCGAATT-TTCCACCTGCTGAATTACTACGCCCCCAAGGAGGATG 120

QY 2447 ATCGCCCTCTCTCTGCTCTTTGAGAAATGTTGGCCATGGCGTTAGTGACAAGAGGG 2506

Db 121 ACCGNCGTCTCTCTGATGTTTGAAGATGTTAGNCTCGTGGCGATTTCGCGAGAGGG 180

QY 2507 ACATCTCGCGATTTCTTGAGTCTAACCCCGGTGATGATGACGCCCAAGAAAGTGTCTGCTG 2566

Db 181 ACATCTCACGTTCTCTGGAGTGAATCCAGTGATGATTTGATGCCATCAAGTTCTTGCTG 240

QY 2567 CACACAGGGCCGTTACTTCTGGGTTAACCTTCTGGCATGACAGGCCCTTTGGCATCCA 2626

Db 241 CTCACAGGGCCCGATCTCTTGGGGCAACCTTACCGGGATGAACAGGCCCGGTATAGCAT 300

QY 2627 CTGTGAATGATAAGCTGGAGCTGCAAGAGTGTCTGGAGCACGGCAGAAATAGCAAGTTCA 2686

Db 301 CAAAGAATGATAAACTCGNGCTGCAGGACTCTCTTGGAAATACAAATAGGATAGCAAGTTAA 360







PR	11-JUL-2000;	2000US-0217487;
PR	11-JUL-2000;	2000US-0217496;
PR	14-JUL-2000;	2000US-0218290;
PR	26-JUL-2000;	2000US-0220963;
PR	26-JUL-2000;	2000US-0220964;
PR	14-AUG-2000;	2000US-0224518;
PR	14-AUG-2000;	2000US-0224519;
PR	14-AUG-2000;	2000US-0225213;
PR	14-AUG-2000;	2000US-0225214;
PR	14-AUG-2000;	2000US-0225266;
PR	14-AUG-2000;	2000US-0225267;
PR	14-AUG-2000;	2000US-0225268;
PR	14-AUG-2000;	2000US-0225270;
PR	14-AUG-2000;	2000US-0225447;
PR	14-AUG-2000;	2000US-0225757;
PR	14-AUG-2000;	2000US-0225758;
PR	14-AUG-2000;	2000US-0225759;
PR	18-AUG-2000;	2000US-0226791;
PR	22-AUG-2000;	2000US-0226681;
PR	22-AUG-2000;	2000US-0226868;
PR	22-AUG-2000;	2000US-0227182;
PR	23-AUG-2000;	2000US-0227009;
PR	30-AUG-2000;	2000US-0228924;
PR	01-SEP-2000;	2000US-0229287;
PR	01-SEP-2000;	2000US-0229343;
PR	01-SEP-2000;	2000US-0229344;
PR	01-SEP-2000;	2000US-0229345;
PR	01-SEP-2000;	2000US-0229509;
PR	05-SEP-2000;	2000US-0229513;
PR	05-SEP-2000;	2000US-0230437;
PR	08-SEP-2000;	2000US-0230438;
PR	08-SEP-2000;	2000US-0230438;
PR	08-SEP-2000;	2000US-0231242;
PR	08-SEP-2000;	2000US-0231243;
PR	08-SEP-2000;	2000US-0231244;
PR	08-SEP-2000;	2000US-0231413;
PR	08-SEP-2000;	2000US-0231414;
PR	08-SEP-2000;	2000US-0232080;
PR	08-SEP-2000;	2000US-0232081;
PR	13-SEP-2000;	2000US-0231968;
PR	14-SEP-2000;	2000US-0232197;
PR	14-SEP-2000;	2000US-0232398;
PR	14-SEP-2000;	2000US-0232399;
PR	14-SEP-2000;	2000US-0232400;
PR	14-SEP-2000;	2000US-0232401;
PR	14-SEP-2000;	2000US-0233063;
PR	14-SEP-2000;	2000US-0233064;
PR	21-SEP-2000;	2000US-0233065;
PR	21-SEP-2000;	2000US-0233423;
PR	21-SEP-2000;	2000US-0236367;
PR	21-SEP-2000;	2000US-0236368;
PR	25-SEP-2000;	2000US-0242787;
PR	25-SEP-2000;	2000US-0242974;
PR	26-SEP-2000;	2000US-0243986;
PR	26-SEP-2000;	2000US-0243984;
PR	27-SEP-2000;	2000US-0235834;
PR	27-SEP-2000;	2000US-0235836;
PR	29-SEP-2000;	2000US-0236327;
PR	29-SEP-2000;	2000US-0236367;
PR	29-SEP-2000;	2000US-0236368;
PR	29-SEP-2000;	2000US-0236369;
PR	29-SEP-2000;	2000US-0236370;
PR	29-SEP-2000;	2000US-0236802;
PR	02-OCT-2000;	2000US-0237037;
PR	02-OCT-2000;	2000US-0237038;
PR	02-OCT-2000;	2000US-0237039;
PR	02-OCT-2000;	2000US-0237040;
PR	13-OCT-2000;	2000US-0239935;
PR	13-OCT-2000;	2000US-0239937;
PR	20-OCT-2000;	2000US-0240960;
PR	20-OCT-2000;	2000US-0241785;
PR	20-OCT-2000;	2000US-0241786;
PR	20-OCT-2000;	2000US-0241787;
PR	20-OCT-2000;	2000US-0241808;
PR	20-OCT-2000;	2000US-0241809;
PR	20-OCT-2000;	2000US-0241826;

PR	20-OCT-2000;	2000US-02422221.
PR	01-NOV-2000;	2000US-0246177.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249300.
PR	17-NOV-2000;	2000US-0249309.
PR	01-DEC-2000;	2000US-0250391.
PR	01-DEC-2000;	2000US-0251160.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-02561479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
XX	(HUMA-)	HUMAN GENOME SCI INC
XX	Rosen CA,	Barash SC, Ruben
XX	WPI;	2001-541565/60.
XX	Nucleic acids encoding 3224 b	
PT	useful for preventing, diagnosing	
PT	cancers and metastases	
PT	-	
XX	Disclosure; SEQ ID NO 12586;	
PS	The invention relates to novel	
XX	(ABBI4678-ABB18001) useful for	
CC	medical conditions e.g. by ph	
CC	isolated from a range of huma	
CC	The nucleic acids, proteins,	
CC	in the diagnosis, treatment,	
CC	and ovarian cancer and other	
CC	marrow, breast, gastrointestinal	
CC	(b) immune disorders e.g. Ad	
CC	haemolytic anaemia, autoimmun	
CC	-	

CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 546 BP; 133 A; 137 C; 171 G; 105 T; 0 other;  
Query Match 4.1%; Score 171.4; DB 22; Length 546;  
Best Local Similarity 81.7%; Pred. No. 1.6e-31;  
Matches 223; Conservative 0; Mismatches 46; Indels 4; Gaps 2;  
DY 384 GGTGGAAGACAGTACACCCCAAGACCCAGCAGTACCAACAGTTCAGCCATGAC 443  
2 GGTGGAAGACAGTACACCCCAAGACCCAGCAGTACCAACAGTTCAGCCATGAC 61  
444 CCAAGACTTGGCCCTCAGATCTGTAACCAATGAGACTTGGAGAGCGAGTAACC 503  
62 CCAAGACTTGGCCCTCAGATCTGTAACCAATGAGACTTGGAGAGCGAGTAACC 121  
DY 504 CCAAGACTTGGAGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 563  
122 CCAAGCGAG 180  
DY 564 GGG---AACTGAGACCCCAAGACCCAGCAGTGTGAGAGAGAGAGAGAGAG 620  
181 GGGTGCAGACTGAGACCCCTGCTGTAAGCTCAAGAGAGAGAGAGAGAGAGAGAG 240  
DY 621 CAAAGAGAGCGCTGAGAGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 653  
241 CAAAGAGAGCGCGAGAGAGCGCGAGCGCGAGAGAGAGAGAGAGAGAGAGAG 273  
Db  
RESULT 14  
AAS86166  
ID AAS86166 standard; cDNA; 2951 BP.  
XX  
AC AAS86166;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #21970.  
XX  
KE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX blood supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PE 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI, 2001-639362/73.  
DR P-PSDB; ABG21979.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnosis, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 1; SEQ ID No 21970; 103bp; English.

XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2951 BP; 632 A; 850 C; 929 G; 540 T; 0 other;  
Query Match 4.1%; Score 169.8; DB 23; Length 2951;  
Best Local Similarity 45.9%; Pred. No. 9.3e-31;  
Matches 837; Conservative 0; Mismatches 952; Indels 36; Gaps 6;  
DY 1133 TTGTGCTTGTGTGATGAGAGCGCGAGCGCGAGCGCTGAAGGACATGCTGGTATGT 1192  
Db 1082 TGTGTCTTGGAGAGCGCGAGCGCGAGCGCGAGCGCTGAAGGACATGCTGGTATGT 1141  
DY 1193 GGTTCGAGATGAGCAAGTCTCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1252  
Db 1142 GGTTCGAGATGAGCAAGTCTCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1201  
DY 1253 TCTGACAGCATTTCCACCGAGCGCGAGCGCGAGCGCGAGCGCGAGCGCGAGCG 1312  
1202 CTGAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1261  
DY 1313 ACGAATCTCTCAGGAGCGCGAGCGCGAGCGCGAGCGCGAGCGCGAGCGCGAGCG 1372  
Db 1262 ACCATGCTCTGAGAGGT-----AACATGAGATGAGGAGAAATGAGGCGCTGAGAGC 1315  
DY 1373 GTGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1432  
Db 1316 AGGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1375  
DY 1433 CCTCGGTGCTTCCAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1492  
Db 1376 CCAACGGGGGCTTCAAGCCACTGAGATGAGAGGCTCAAAACCAACAGCGAACAG 1435  
DY 1493 ATCCTTCAAGGAATTTACCGAGCATGTGGGTGAGCGCTGAAGCGCTGTTAGGCC 1552  
Db 1436 GTGGGAATGAGTCCCATGAGCAGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1495  
DY 1553 CACCCGACGAGCAAGAAACCCAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1612  
Db 1496 GAGTCC-----GCTCAAGACAAATTCCTATTAACAACGGCAAGAGCGAGATGAAGA 1550  
DY 1613 TCATTGATGAGCGCAAGAGAGAGCGCTGTGTATGAGGTGCGCAGAAAGTGCAGAAACA 1672  
Db 1551 TCAGAGCGGAGGTGATTTGTGGTACCTGGGATCATGGACAGATGGAGAGAGAGCGCTG 1610  
DY 1673 TCAGAGCATTTGATATCATGATGAGAGCGCTCAATGTCAACCTGAGAGAGAGCGCTTCA 1732  
Db 1611 CAGATCGAAGTAATGATGAT-----CCCCCTCACACCTGGGCTAGAGCGCTTACGCC 1662  
DY 1733 TTGAGAGCATGTGCAGAACTGTAAGAACTGCTTGTGAGTGTGCTTACAGATATGACG 1792  
Db 1663 TGCTTGTCTGTCAACAATGCGGGTAAAGTCTCTACTACTGCTTGAACCTTCTCTCC- 1721

QY 1793 ACGATGGGTACACAGTCTTATGTCACCATCTGCTGTGGGGCGGTGAAGTGTCTATGTGTG 1852  
Db 1722 -CCTTGGCTCTTAACCTCCCTCTCTTCCCTCGAGGGCCGAGAGCTGTCTGCTTTGCA 1780  
QY 1853 GGAACAACACTGCTCGAGTGTCTTTGTGTCGAGTGTGTGATCTCTTGTGTGGGGCCAG 1912  
Db 1781 GCAACACAGAGTGTCTCCGGTGTAGCACTGGCCCTGTGGGTGTGATGTGGTGGCCCCC 1840  
QY 1913 GAGCTGCTCAGGAGCCATTAAGGAAGACCCCTGGAACTGTCTAATGTGCGGGCATAAAG 1972  
Db 1841 AAGGCTCTACGTTCTTGCACTTAGGAGCCCTGGAGCTGTATACATGTCTCCCGCAGC 1900  
QY 1973 GCACCTATGGGTGCTGCGAAGACGGGAAGACTGGCCCTTCTCGACTCCAGATGTCTTTG 2032  
Db 1901 GCTGTATGGCGTCTGCGGCGCGGAGACTGGACGTGGCCCTGCAGGCCCTTCTCA 1960  
QY 2033 CCAATAACCATGACACAGGAATTTGACCCCCCAAGGTTTACCCACCTGTGCCAGCTGAGA 2092  
Db 1961 CCAGTGACACGGGGCTTGAATATGTAAAGCCACAGGCTCCCGCCCTTACCCACAGATCC 2020  
QY 2093 AGAGGAAGCCATCCCGCTGCTGTCTCTTTGATGGATTGTACAGGGCTCTCTGTGTGC 2152  
Db 2021 CAGGGGCACAGGTGTGGAAGCTCTGTTTGTATGTCATCGCACAGGTGAGTTGCGGG 2080  
QY 2153 TGAAGCACTGGGCATCCAGTGGACCGCTACATTGCTCCGAGGTGTGTAGGACTCCA 2212  
Db 2081 AACACCTGGAGACATGCTATCGTGTACAAACAGGGTAGCCAGGGAGTCAGAAGAGTCCA 2140  
QY 2213 TCACGCTGGGCATGTGCGGCACACAGGAAAGATCATGTACGTGCGGGACGTCGCGAGCG 2272  
Db 2141 TTGCTGTGGAACCGTGAAGCAGCAGGGGAATATCAAAATACGTGAACGACGTGAGGAACA 2200  
QY 2273 TCACACAGAGCATATCCAGGAGTGGGCCCATTCGACCTGGTGTGAGGCGACCTCCCT 2332  
Db 2201 TCACAAAGAAATGTAGGGCAGTGTATACCTTCGCGGCCCTCATCTCTTCGTCTTTT 2260  
QY 2333 GCAATGACCTCTCAATTGTCAACCTGCCGCAAGGACTTTATG-----AGGTAATG 2386  
Db 2261 TCCCGAGTCTCCACACCCGAAATCCAGCCAGGAAGGCTGTATGTGAGCATCTTCTC 2320  
QY 2387 GCGGCTCTCTTTGAGTCTACCGCTCTG-----CATGATCGCGGCCCAAGG 2437  
Db 2321 TCTGGCAGTCCCTGGAGAGCTATGTACCTGACCACCTGGCCAGGTGAGGAATAAGG 2380  
QY 2438 AGGAGATGATGCCCTCTTCTTGGCTCTTTGAGATGTGGTGGCCATGGCGGTAGTG 2497  
Db 2381 TGGGTGGCTGGAGTTTCAAAATGAACCTCGCGCTGTCTCTTTCTGAGCACAGAGG 2440  
QY 2498 ACAAGAGGACATCTCGCATTTCTTGAGTCTAACCCCGTGTATGATTCACGCCAAAGAG 2557  
Db 2441 TACAGAGGACATCTACGGTTCTCGAGGTGAGGAACTGCGGGACCTGATTGTCAAG 2500  
QY 2558 TGTCTGTGCACACAGGGCCCGTTACTTCTGGGGTAACCTTCTTGGCATGACAGCCCTT 2617  
Db 2501 ACAGCCAGGGCAGGAAAGCGCTGCTGGCAGTGAACCTACCCGGGATGAACAGGTAA 2560  
QY 2618 TGGCATCCACTGTGAATGATAAGCTGGAGTGCAGAGTGTCTGGAGCACGGCAGAAATAG 2677  
Db 2561 AAAGGCTCTTAGTGGTCAAGTAACAGCCAAAGTTAAATATGTGATAACAAGCTCTGACA 2620  
QY 2678 CCAAGTTCAGAAAGTGAAGACCATTTACACAGGTCAAACTCTATAAGCAGGCGCAAG 2737  
Db 2621 CCAAGTAAGACGAGCTGTGGCCCTCTGGAAATATGCACTTGGTGAACCTCAAGTGGGA 2680  
QY 2738 ACCAGCATTTCCCGCTTCTCATGAACGAGAGGAGCATCTCTGTGGTGCATGAAATGG 2797  
Db 2681 CTTGGAGATCCCTGTTGTATGAATGGCAAGAGATGTTTGTGTGTCATGAGCTGAGCTG 2740  
QY 2798 AAAGGTGTTTGGCTTCCCGCTCCACTACACAGAGCTCTCAACATGAGCCCTTGGGCA 2857  
Db 2741 AAAGGTGACAGGCTGCACTTGGAGAGGGAATCTGTGTAGATCAAAACAAATGGGCA 2800  
QY 2858 GGCAGAGACTGTGGGCCGATCTGTGGAGCGTGCCTGATCCGCCACCTCTTCGCTCCGC 2917

Db 2801 GACATGGGCTGCTGGGAAGTCTCGAGCGTCTGTCTATCGACACCTCTTCGCCCTC 2860  
QY 2918 TGAAGGAATATTTTGTCTGTGTGTGA 2942  
Db 2861 TGAAGGACTACTTTTGCATGTGAATA 2885  
RESULT 15  
ABAI3731  
ID ABAI3731 standard; cDNA; 283 BP.  
XX ABAI3731;  
AC ABAI3731;  
XX  
DT 23-JAN-2002 (first entry)  
XX  
DE Human nervous system related polynucleotide SEQ ID NO 2738.  
XX  
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;  
KW antiparkinsonian; antiskilling; antianemic; antiarthritic; cancer;  
KW antirheumatic; hepatotropic; cerebrotective; antiinflammatory;  
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200159063-A2.  
XX PD 16-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01334.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180828.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226275.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.



Db 141 AATGGGACTTGGAGAGCGAGTGAAGCCCGAGCCAGAGGAGGGAGCCCTGCTGGGGG 200  
QY 535 CAGAGGGTGGGGCCCCAGCTGAAGGAGAGGG---AACTGAGACCCACAGAGCCTCC 591  
Db 201 CAGAGGGCGGGCCCCAGAGGAGAGGGTGCAGCTGAGACCCCTGCTGAAGCCTCA 260  
QY 592 AGAGCTGTGAGAAATGGCTGCTG 614  
Db 261 AGAGCAGTGGAAAAATGGCTGCTG 283

Search completed: November 22, 2002, 09:09:32  
Job time : 605.945 secs

4  
4  
4  
4  
4  
4  
4  
4  
4

D5 1169 AAGCTGGTTTCATATGGAACCCAGTACCACATCTCGAAGAACCAAGCAGG

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12

**THIS PAGE BLANK (USPTO)**

```

OY 321 G1yLysThrPheSerSerSerProGlyGlySerLeuGluAspGluLeuLysProMetLeu 340
Db 1229 GGCAAGACCTTCTCCAGCAGCTCTGGAGAGTCACTGGAGACCACTGAAGCCCATGCTG 1288
OY 341 G1uTrpAlaHisGlyGlyPheLysProThrGlyIleGluGlyLeuLysProAsnLysLys 360
Db 1289 GAGTGGGCGCCAGCGTGGCTTCAGACCTACTGGGATGAGGCGCTCAAAACCAAGAG 1348
OY 361 G1nProValIleAsnLysSerLysValArgArgSerSerSerAsnLeuGluProArg 380
Db 1349 CAACCACTGGTTAAATAGTCGAGAGGTCGCTTCAGACAGTGAAGCTTGAAGCCAGG 1408
OY 381 ArgArgGluAsnLysSerArgArgArgThrThrAsnAspSerAlaIleSerGlySerPro 400
Db 1409 AGACCGAGAGAACAAAGTCGAGAGCGCAACCAATGACTCTGCTGCTTCTGAGTCTCC 1468
OY ProProLysArgLeuLysThrAsnSerTyrglyGlyLysAspArgGlyGluAspGluGlu 420
Db CCACCCAAAGCGCTCAAGACAAATAGCTATGGCGGAGAGACCGAGGGGAGATGAGAG 1528
OY 421 SerArgGluArgMetAlaSerGlyValThrAsnAsnLysGlyAsnLeuGluAspArgCys 440
Db 1529 AGCCGAGAGACCGATGCTTCTGAAGTCAACCAACAGGAGCAATCTGAGAGCCGCTGT 1588
OY 441 LeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyLysCys 460
Db 1589 TTGTCTGTGGAAAGAGAACCTCTGTCTTCCACCCCTCTTGAAGGTGGGCTCTGT 1648
OY 461 GlnSerCysArgAspArgPheLeuGluLeuPheTyrglyMetTyrglyAspGluAspGlyTyrgln 480
Db 1649 CAGAGTGGCCGGGATCGCTTCTAGAGCTCTTCAATGATATGATGAGAGCGGCTATCG 1708
OY 481 SerTyrglyThrValCysCysGluGlyArgGluLeuLeuLysSerAsnThrSerCys 500
Db 1709 TCCTACTGCACCGTGTCTGTAGAGGCGCTGAGACTGCTGTGCGAGTAAACAAAGCTGC 1768
OY 501 CysArgCysPheCysValGluCysLeuGluValLeuValGlyAlaGlyThrAlaGluAsp 520
Db 1769 TGCAGATGCTTCTGTGTGGAGTGTCTGAGAGTGTGTGGGCGCAGACAGCTGAGCAT 1828
OY 521 AlaLysLeuGluGluProTrpSerCysTyrglyMetCysLeuProGluArgCysHisGlyVal 540
Db 1829 GCCAAGCTGCAGAGAACCTGAGCTGTATATGTGCTCTCCAGCGCTGCCATGGGGTTC 1888
OY LeuArgArgArgLysAspTrpAsnMetArgLeuGluAspPhePheThrThrAspProAsp 560
Db CTCCGACCGCAGAGAAATGGAACATGCGCTGCAAGACTTCTTCACTACTGCTCTGAC 1948
OY 561 LeuGluGluPheGluProProLysLeuTyrglyProAlaIleProAlaIleLysArgArgPro 580
Db 1949 CTGGAGAAATTTGAGCCACCAAGTTGTACCAGCAATTCCTGCAAGCCAAAGAGGCC 2008
OY 581 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyTyrglyLeuValLeuLysGluLeu 600
Db 2009 ATTAGAGTCTGTCTCTTGTGATGCAATTCGCAAGCGGGTACTTGTGCTCAAGAGTTG 2068
OY 601 GlyIleLysValGluLysTyrglyIleAlaSerGlyValCysAlaGluSerIleAlaValGly 620
Db 2069 GGATTTAAAGTGAAGAAATGACTTGTCTCGAAAGTCTGTGCAAGTCCATCGCTGTGGGA 2128
OY 621 ThrValLysHisGluGlyGlnIleLysTyrglyValAsnAspValArgLysIleThrLysLys 640
Db 2129 ACTGTTAAAGCATGAGCGCCAGATCAAAATATGTCATGACGTCCGGAATAACCAAGAAA 2188
OY 641 AsnIleGluGluTrpGlyProPheAspLeuValIleGlyLysSerProCysAsnAspLeu 660
Db 2189 AATATTGAAGAGTGGGCGCTTGTGACTTGTGATTTGGTGAAGCCCATGCAATGATCTC 2248
OY 661 SerAsnValAsnProAlaArgLysGlyLeuTyrglyGlyTyrglyArgLeuPhePheGlu 680
Db 2249 TCTAACCTCATCTCGCCGCAAGGTTATATAGGGCAGCAGAGAGGCTCTTCTTGAG 2308
OY 681 PheTyrglyLeuLeuAsnTyrglyThrArgProLysGluGlyAspAsnArgProPhePheTrp 700

```

```

Db 2309 TTTTACCACTTCTGATTTATACCCGCCCCCAAGAGGCGCACAACCGTCCATTCTTCTGG 2368
OY 701 MetPheGluAsnValValAlaMetLysValAsnAspLysAspIleSerArgPheLeu 720
Db 2369 ATGTTTCAGAAATGTTGTGGCCATGAAGTGAATGACAGAAAGACATCTCAAGATTCTCTG 2428
OY 721 AlaCysAsnProValMetIleAspAlaIleLysValSerAlaAlaHisArgAlaArgTyrg 740
Db 2429 GCATGTAAACCAAGATGATGCATGCCATCAAGGTGTGCTGCTCAGAGGGCCCGGTAC 2488
OY 741 PheTrpGlyAsnLeuProGlyMetAsnArgProValMetAlaSerLysAsnAspLysLeu 760
Db 2489 TTCTGGGATTAACCTAACCGGATGAACAGGCCCCCTGATGGCTTCAAGAAATGATAGCTC 2548
OY 761 GluLeuGluAspCysLeuGluPheSerArgThrAlaLysLeuLysValGlnThrIle 780
Db 2549 GAGCTGCAGAGACTGCTGAGCTTCACTAGTACGACGCAAGTTAAAGAAAGTGCAGACATA 2608
OY 781 ThrThrLysSerAsnSerIleArgGlnGlyLysAsnGluLeuPheProValValMetAsn 800
Db 2609 ACCACCAAGTCAAACTCCATCAGACAGGCGCAAAACCGCTTTCCTGTAGTCATGAT 2668
OY 801 GlyLysAspAspValLeuTrpCysThrGluLeuGluArgIlePheGlyPheProAlaHis 820
Db 2669 GGCAGAGACGACGTTTGTGTGTCATGAGCTCGAAAGGATCTTGGCTTCCCTGCTCAC 2728
OY 821 TyrglyAspValSerAsnMetGlyArgGlyAlaArgGluLysLeuGluGlyArgSerTrp 840
Db 2729 TACAGGACGCTGTCCAAATGAGGCGCGCGCCCTCAGAACCTGTGGGCGAGCTCTGG 2788
OY 841 SerValProValIleArgHisLeuPheAlaProLeuLysAspTyrglyPheAlaCysGlu 859
Db 2789 AGTGTACCGGTCAACAGACCTGTTGCCCTTGAAGAGACTTGTGCTGTGCA 2845

```

RESULT 2  
AF151973  
LOCUS



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 22, 2002, 03:58:32 ; Search time 4894.5 Seconds  
(without alignments)  
5107.633 Million cell updates/sec

Title: US-09-720-086-6  
Perfect score: 4590  
Sequence: 1 MKGDSRHLHBEAGSGYECC.....MSVPVIRHLFAPLKDYFACE 859

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool/US09720086/runat.18112002.092832.22195/app.query.fasta.1.4252  
-DB=GenBdb1 -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pico -NORM=ext -HEAFSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09720086.GCCN.1.11308.@runat.18112002.092832.22195 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOB=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenBdb1:  
1: gb\_ba:\*  
2: gb\_hlg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_ats:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_ats:\*  
28: em\_un:\*

29: em\_vi:\*  
30: em\_hlg\_hum:\*  
31: em\_hlg\_inv:\*  
32: em\_hlg\_other:\*  
33: em\_hlg\_mus:\*  
34: em\_hlg\_pin:\*  
35: em\_hlg\_rod:\*  
36: em\_hlg\_mam:\*  
37: em\_hlg\_vrt:\*  
38: em\_sy:\*  
39: em\_hlgo\_hum:\*  
40: em\_hlgo\_mus:\*  
41: em\_hlgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4590	100.0	4195	10 AF068626	AF068626 Mus muscu
2	4573	99.6	4223	10 AF151973	AF151973 Mus muscu
3	4573	99.6	4338	10 AF151969	AF151969 Mus muscu
4	4472	97.4	4135	10 AF068627	AF068627 Mus muscu
5	4455	97.1	4163	10 AF151974	AF151974 Mus muscu
6	4455	97.1	4278	10 AF151970	AF151970 Mus muscu
7	4223.5	92.0	4006	10 AY078427	AY078427 Mus muscu
8	4206.5	91.6	4034	10 AF151975	AF151975 Mus muscu
9	4206.5	91.6	4149	10 AF151972	AF151972 Mus muscu
10	4105.5	89.4	3946	10 AF068628	AF068628 Mus muscu
11	4088.5	89.1	3974	10 AF151976	AF151976 Mus muscu
12	4088.5	89.1	4089	10 AF151971	AF151971 Mus muscu
13	3759	81.9	4145	9 AF156488	AF156488 Homo sapi
14	3759	81.9	4335	9 AF331857	AF331857 Homo sapi
15	3675	80.1	4267	9 AF176228	AF176228 Homo sapi
16	3331.5	72.6	3897	9 AF156487	AF156487 Homo sapi
17	2453	53.4	3017	9 AK001191	AK001191 Homo sapi
18	2206.5	48.1	3005	9 AF067972	AF067972 Homo sapi
19	2206.5	48.1	4258	9 AF331856	AF331856 Homo sapi
20	2185	47.6	4192	10 AF068625	AF068625 Mus muscu
21	2184	47.6	4094	10 BC007466	BC007466 Mus muscu
22	1822.5	39.7	2848	5 AF135438	AF135438 Dario rer
23	1734.5	37.8	2191	9 AK025230	AK025230 Homo sapi
24	1428.5	31.1	176697	2 AC112586	AC112586 Rattus no
25	1352	29.5	123936	2 AC120824	AC120824 Rattus no
26	1271	27.7	167568	2 AC111734	AC111734 Rattus no
27	1216	26.5	244329	2 AC107644	AC107644 Mus muscu
28	1209	26.3	1758	9 BC018214	BC018214 Homo sapi
29	1183	25.8	168651	9 AC009474	AC009474 Homo sapi
30	1128	24.6	2127	9 AF129267	AF129267 Homo sapi
31	987.5	21.5	2057	9 AF129268	AF129268 Homo sapi
32	882	19.2	2008	9 AF129269	AF129269 Homo sapi
33	862	18.8	2077	6 AR129189	AR129189 Sequence
34	813	17.7	225045	2 AL833803	AL833803 Mus muscu
35	811.5	17.7	249345	2 AC122356	AC122356 Mus muscu
36	695.5	15.2	188936	2 AC128509	AC128509 Rattus no
37	675.5	14.7	1705	9 AF194032	AF194032 Homo sapi
38	673	14.7	1397	9 BC002560	BC002560 Homo sapi
39	660	14.4	1589	10 AF220524	AF220524 Mus muscu
40	660	14.4	1645	10 MMU304467	MMU304467 Mus muscu
41	632.5	13.8	110000	2 AL354832	AL354832 Homo sapi
42	632.5	13.8	118899	9 HS1085F17	HS1085F17 Human DNA
43	618	13.5	210269	2 AC116459	AC116459 Mus muscu
44	597.5	13.0	119630	2 AC112040	AC112040 Rattus no
45	571	12.4	153533	2 AC095199	AC095199 Rattus no

RESULT 1

ALIGNMENTS

Mon Nov 25 08:18:28 2002

```

AF068626
LOCUS       AF068626             4195 bp      mRNA      linear      ROD 06-DEC-1999
DEFINITION Mus musculus DNA cytosine-5 methyltransferase 3B1 (Dnmt3b) mRNA,
alternatively spliced, complete cds.
ACCESSION   AF068626
VERSION     AF068626.2
KEYWORDS    GI:6449469
SOURCE      Mus musculus.
  ORGANISM  Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 4195)
AUTHORS    Okano, M., Xie, S. and Li, E.
TITLE      Cloning and characterization of a family of novel mammalian DNA
            (cytosine-5) methyltransferases
JOURNAL    Nat. Genet. 19 (3), 219-220 (1998)
MEDLINE    98324766
PUBMED     9662389
REFERENCE   2 (bases 1 to 4195)
AUTHORS    Xie, S., Okano, M. and Li, E.
TITLE      Direct Submission
JOURNAL    Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,
            Charlestown, MA 02129, USA
REFERENCE   3 (bases 1 to 4195)
AUTHORS    Okano, M., Chijiwa, T., Sasaki, H. and Li, E.
TITLE      Direct Submission
JOURNAL    Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,
            Charlestown, MA 02129, USA
REMARK     Sequence update by submitter
COMMENT    On Nov 18, 1999 this sequence version replaced gi:3327979.
FEATURES   Location/Qualifiers
            1..4195
                /organism="Mus musculus"
                /db_xref="taxon:10090"
                /chromosome="2"
                /map="84.0 CM"
            1..4195
                /gene="Dnmt3b"
            1..4195
                /gene="Dnmt3b"
            1..4195
                /note="similar to EST sequences deposited in GenBank
                Accession Numbers AA116634, AA119979, AA177277, AA210568,
                AA407106, and AA575617"
            269..2848
                /gene="Dnmt3b"
                /function="de novo DNA methylation"
                /note="alternatively spliced product; contains Cys-rich
                region; C-terminal region is similar to corresponding
                region of DNA cytosine-5 methyltransferase 3A"
            /codon_start=1
            /product="DNA cytosine-5 methyltransferase 3B1"
            /protein_id="AAC40178.2"
            /db_xref="GI:6449470"
            /translation="MKGDSRHLNBEAGSGYECIIVNGNFSDQSDTKDAPSPVLE
            AICTPEVCTPETRGRSSRLSKREVSSLLNYTQDMTGDGDDDEVDGNGSDILMEK
            LTRFTRTRTRSPAVTRHNGTSSLERQASPRITRGRGRHHVQEVFPFPAIR
            SRRRASASSTPSPASVDFMEVTPKSVSTPSVDLSQDQDQEGMDTTOVDAESRD
            GDSTYQDDKEFGIGDLVWGKIKGFSWPMVSNKATSKRQAMPGRWVQWFDGKRF
            SEISADKLVALGLFQHLNFKLVSVKAMVHTLEKARVAGKTFSSSPGESLED
            OLKPMLEAHGKFKPTIEGLKPKQPVVNVKSRSDSRNLEPRRNRKSRRTTN
            DSAASEPPPKLNTNSVGGKDGDEESRERMASEVTNNKGNLEDCLSCGKKNPVS
            FPLPREGGLQSCQRDRFLPFLYMDGQVQSVCTVCEGRILLCSNCCRCPCVPC
            LEVLVAGTAGDAKLQEPWSCMCLFQRCHVLRKRRDNWRLQDPTTDPDLBEFEP
            PKLYPAI PAARRPRIVLSLFDGIATGILVLKELGKIKVKIASEVCAESIAVGTVKH
            EQQIVYNDVRKTKNIKEWGPFDLVIGSPCNDLSNPNPARKMIDGEGGRFLFFFY
            HLLNTRPKGDNRPFFWFMFNVAMKNDKDISRLACNPNMIDAIKVSAAHRRY
            FGNLPGMNRPVMASKNKLEQDCLFSTRKLVKQVITTTKNSIRQGNQIPLFV
            MNGKDVLCWLTLELRFIFGPAHYTDVSNMGRGARQKLLGRSWSVPVIRHFLPKDYF
            ACE"
BASE COUNT      1061 a 1059 c 1125 g 950 t
ORIGIN
Alignment Scores:

```

```

Pred. No.: 2.1e-249      Length: 4195
Score: 4590.00      Matches: 859
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 100.00%      Indels: 0
DB: 10      Gaps: 0

US-09-720-086-6 (1-859) x AF068626 (1-4195)

QY 1 MetLysGlyAspSerArgHisLeuAsnGluGluGlyAlaSerGlyTyrGluGluCys 20
Db 269 ATGAAGGAGAGACAGCAGACATCTGAATGAAGAAGAGGGTCCAGCGGTATGAGGATGC 328
QY 21 IleIleValAsnGlyAsnPheSerAspGlnSerSerAspThrLysAspAlaProSerPro 40
Db 329 ATTTATCGTTAATGGAACTTCAGTCACAGTCTCAGACACAGGAGGATGCTCCCTCACCC 388
QY 41 ProValLeuGluAlaIleCysThrGluProValCysThrProGluThrArgGlyArg 60
Db 389 CCAGTCTTTGGAGGCAATCTGCACAGAGCCAGTCTGCACACAGGACCCAGAGCCGAGG 448
QY 61 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuAsnTyrThrGlnAspMet 80
Db 449 TCAAGTCCCGCTCTCTAAGAGGGAGGTCTCCAGCTTCTGAATTTACACGAGGACATG 508
QY 81 ThrGlyAspGlyAspArgAspAspGluValAspAspGlyAsnGlySerAspIleLeuMet 100
Db 509 ACAGAGATGGAGACAGATGATGAAGTAGATGATGGGATGCTGATATTTCTAATG 568
QY 101 ProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSerProAlaValArg 120
Db 569 CCAAGCTCACCCGTGAGACCAAGGACACACAGGACGCTCTGAAAGCCCGGCTGCCGA 628
QY 121 ThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnArgAlaSerProArgIleThr 140
Db 629 ACCCGACATAGCAATGGGACCTCCAGCTTGGAGAGCAAGAGAGCTCCCCAGATCACC 688
QY 141 ArgGlyArgGlnGlyArgHisHisValGlnGluTyrProValGluPheProAlaThrArg 160
Db 689 CGAGTCCGACAGGCGCCGCCCATGTGCAGAGGATACCTGTGGAGTTTCGGGTACCAGG 748
QY 161 SerArgArgArgArgAlaSerSerSerAlaSerThrProTrpSerSerProAlaSerVal 180
Db 749 TCTCGGAGACGTCGAGCATCGTCTTTCAGCAAGCACGCGCATCCCTGCCAGCGTC 808
QY 181 AppPheMetGluGluValThrProLysSerValSerThrProSerValAspLeuSerGln 200
Db 809 GACTTCATGGAAAGAGTGACACCTTAAGAGCGTCAGTACCCCATCAGTTGACTTGAGCCAG 868
QY 201 AspGlyAspGlnGluGlyMetAspThrThrGlnValAspAlaGluSerArgAspGlyAsp 220
Db 869 GATGGAGATCAGAGGGGTATGGATACACACAGAGTGGATGCAGAGAGCAGAGATGGAGAC 928
QY 221 SerThrGluTyrGlnAspAspLysGluPheGlyIleGlyAspLeuValTrpGlyLysIle 240
Db 929 AGCAGAGATATCAGGATGATTAAGAGATTGGATAGGTAGTACCTCGTGGGGAAGATC 988
QY 241 LysGlyPheSerTrpTrpProAlaMetValValSerTrpLysAlaThrSerLysArgGln 260
Db 989 AAGGGCTTCTCTGGTGGGCTGCGCATGGTGGTGTCTCTGGAAAGACCTCCCAAGGCAG 1048
QY 261 AlaMetProGlyMetArgTrpValGlnTrpPheGlyAspGlyLysPheSerGluIleSer 280
Db 1049 GCCATGCCCGGAATCGCTGGGTACAGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGT 1108
QY 281 AlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsn 300
Db 1109 GCTGCAAACTGGTGGCTCTGGGGCTGTTCCAGCCAGACACTTAATCTGCTACCTCAAT 1168
QY 301 LysLeuValSerTyrArgLysAlaMetTyrHisThrLeuGluLysAlaArgValArgAla 320
Db 1169 AAGCTGGTTTCTTATAGGAAGGCCATGTACCACTCTGGGAAAGCCAGGGTTCGAGCT 1228

```

QY 321 GlyLysThrPheSerSerSerProGlyGluSerLeuGluuAspGluLeuLysProMetLeu 340  
 DB 1229 GGCAGAGACCTTCTCCACAGATCTGGAGATCCTGGAGACAGCTGAAAGCCCATGCTG 1288  
 QY 341 GluTrpAlaHisGlyGlyPheLysProThrGlyLysGluGlyLeuLysProAsnLysLys 360  
 DB 1289 GAGTGGGCCCCACGGTGGCTTAAAGCTTACTGGATCGAGGCTCAAAACCAACAGAG 1348  
 QY 361 GluProValValAsnLysSerLysValArgArgSerAspSerArgAsnLeuGluProArg 380  
 DB 1349 CAACCAAGTGTAAATAGTCCAGAGTGCCTGCTTCAAGACAGTAGGAACCTTGAAGCC 1408  
 QY 381 ArgArgGluAsnLysSerArgArgArgThrThrAsnAspSerAlaLysArgLysPro 400  
 DB 1409 AGAGCGGAGAACAAAGATCGAAGACCAACCAATGACTGCTGCTTCTGAGTCCCCC 1468  
 QY 401 ProProLysArgLeuLysThrAsnSerTyrGlyLysAspArgGlyLysAspGluGlu 420  
 DB 1469 CCAACCAAGCGCTCAAGACAAATAGCTATGGCGGGAAGAGCCGAGGCGAGTGAAGAG 1528  
 QY 421 SerArgGluArgMetAlaSerGlyValThrAsnAsnLysGlyAsnLeuGluAspArgCys 440  
 DB 1529 AGCCGAGAACGGATGGCTTCTGAGATCACCAACAAAGGCAATCTGGAAAGCCGCTGT 1588  
 QY 441 LeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyLysCys 460  
 DB 1589 TTGTCTCTGGGAAAGAAAGAACCTGTCTTCAACCCCTCTTGAAGGCTGGCTCTGT 1648  
 QY 461 GluSerCysArgAspArgPheLeuGluLeuPheTyrMetTyrAspGluAspGlyTyrGln 480  
 DB 1649 CAGAGTGGCGGAGATGCTTCTAGAGCTTCTTCAATGATGATGAGAGCGGATATAG 1708  
 QY 481 SerTyrCysThrValCysCysGluGlyArgGluLeuLeuLeuLeuCysSerAsnThrSerCys 500  
 DB 1709 TCCTACTGACCTGTCTGTGAGGCGCGTGAACCTGCTGCTGTGCACTAACACAAAGCTGC 1768  
 QY 501 CysArgCysPheCysValGluCysLeuGluValLeuValGlyLysThrAlaGluAsp 520  
 DB 1769 TGCAGATGCTTCTGTGTGGAGATGCTGAGAGTCTGCTGCGGCGAGGACAGCTGAGAT 1828  
 QY 521 AlaLysLeuGluGluProTyrPheCysTyrMetCysLeuProGluArgCysHisGlyVal 540  
 DB 1829 GCCAAGCTCGAGAACCTGTGAGCTGCTATATGTCCTCCCTCAGCGCTCCATGGGCTC 1888  
 QY 541 LeuArgArgArgLysAspTyrAsnMetArgLeuGluAspPheThrThrAspProAsp 560  
 DB 1889 CTCGAGCGAGAAAGATTGGAACAATGCGCTGGAAGCTTCTTCACTACTGATCCTTAC 1948  
 QY 561 LeuGluGluPheGluProProLysLeuTyrProAlaLeuProAlaAlaLysArgArgPro 580  
 DB 1949 CTGGAAGATTGAGCCACCAAGTTGTACCCAGCAATTCCTGCAAGCCAAAGAGGCC 2008  
 QY 581 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeu 600  
 DB 2009 ATTAGAGTCTGCTCTGTTGATGGAATTGCAACGGGTAATTTGTCCTCAAGAGATTG 2068  
 QY 601 GlyIleLysValGluLysTyrIleAlaSerGluValCysAlaGluSerIleAlaValGly 620  
 DB 2069 GGTATTTAAAGTGAAGAAAGTACATTCGCTCGAAGTCTGTGAGAGTCCATCGCTGTGGA 2128  
 QY 621 ThrValLysHisGluGlyGlnIleLysTyrValAsnAspValArgLysIleThrLysLys 640  
 DB 2129 ACTGTTAAACATGAAAGCCAGATCAATATGTCAATGACGTCCGAAAAATCACCAGAAA 2188  
 QY 641 AsnIleGluGluTrpGlyProPheAspLeuValIleGlyLysSerProCysAsnAspLeu 660  
 DB 2189 AATATTGAAGAGTGGGCGCCGCTTGCACCTGGTGTGATGGTGAAGCCCATCATGATCTC 2248  
 QY 661 SerAsnValAsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPhePheGlu 680  
 DB 2249 TCTAACGTCAATCTGCGCCGCAAGGCTTTATAGAGGCGACAGAAAGGCTCTTCTCGAG 2308  
 QY 681 PheTyrHisLeuLeuAsnTyrThrArgProLysGluGlyAspAsnArgProPhePheTrp 700

DB 2309 TTTTACCACTTCTGTAATTTATACCCGCCCCAAAGAGGCGACAAACCGTTCATTTCTGG 2368  
 QY 701 MetPheGluAsnValValAlaMetLysValAsnAspLysAspIleSerArgPheLeu 720  
 DB 2369 ATGTTCCGAGAAATGTTTGGCCATGAAGTGAATGACAAAGAAAGATCTCAAGATTCTTG 2428  
 QY 721 AlaCysAsnProValMetIleAspAlaIleLysValSerAlaAlaHisArgLysTyr 740  
 DB 2429 GCATGTAAACCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2488  
 QY 741 PheTrpGlyAsnLeuProGlyMetLysAsnArgProValMetAlaSerLysAsnAspLeu 760  
 DB 2489 TTCTGGGGTAACCTTACCCGGAATGAACAGGCGCCGTGATGCTTCAAGAAATGATTAAGTCTC 2548  
 QY 761 GluLeuGluAspCysLeuGluPheSerArgThrAlaLysLeuLysLysValGlnThrIle 780  
 DB 2549 GAGCTGCAGAGATCGCTCGAGTTCAGTAGAGACAGCAAAAGTTAAAGAAAGTGCAGACATA 2608  
 QY 781 ThrThrLysSerAsnSerIleArgGlnGlyLysAsnGluLeuPheProValMetAsn 800  
 DB 2609 ACCACCAAGTGCAGATCCATCCAGACAGGCGCAAAACAGCTTTCCCTGATGATGATGAT 2668  
 QY 801 GlyLysAspAspValLeuTrpCysThrGluLeuGluArgIlePheGlyPheProAlaHis 820  
 DB 2669 GGCAAGACGACGTTTGTGTGTGCTGACATGAGCTGAAAGCATCTTCCGCTCCCTGCTCAC 2728  
 QY 821 TyrThrAspValSerAsnMetGlyArgGlyAlaArgGluLysLeuLeuGluLysSerTyr 840  
 DB 2729 TACACGAGCTGTCCAAACAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2788  
 QY 841 SerValProValIleArgHisLeuPheAlaProLeuLysAspTyrPheAlaCysGlu 859  
 DB 2789 AGTGTACCGGTATCATGACACCTGTTGCGCCCTTGAAGAGCATCTTCCCTTGAA 2845  
 RESULT 2  
 AFI51973 4223 bp mRNA linear ROD 08-JUN-2000  
 LOCUS AFI51973  
 DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 5  
 (Dnmt3b) mRNA, complete cds.  
 ACCESSION AFI51973  
 VERSION AFI51973.1 GI:8347127  
 KEYWORDS  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 REFERENCE 1 (bases 1 to 4223)  
 Ylin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and Shen,Y.  
 Cloning of full-length Dnmt3b cDNA and its alternative splicing isoforms in mouse embryonic tissue  
 Unpublished  
 2 (bases 1 to 4223)  
 Ylin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and Shen,Y.  
 Direct Submision  
 Submitted (17-MAY-1999) Department of Biochemistry and Molecular Biology,  
 Institution of Basic Medical Sciences, Chinese Academy of Medical Sciences,  
 Dong Dan San Tiao 5, Beijing City 100005.  
 P.R.China  
 FEATURES  
 source  
 1..4223  
 /organism="Mus musculus"  
 /strain="KM"  
 /db\_xref="taxon:10090"  
 /dev\_stage="8-9 day old embryo"  
 1..4223  
 /gene="Dnmt3b"  
 297..2876  
 /gene="Dnmt3b"  
 /function="methylates cytosine in DNA"  
 /note="alternatively spliced"

```
/codon_start=1
/produCt="DNA cytosine-specific methyltransferase isoform
5"
/protein_id="AAF74519.1"
/db_xref="GI:8347128"
/translation="MKGDSRLHNEBEGASGYBECIIVNGNFSQSDTKDAPPPVLE
AICTEPCVTPETRRSSRLSKRSYLLNYTQDMTGDGDDDEVDVNGSDILMPK
LRTREKTRTRSESAPVTRHSNGTSSLERQASPRITRGRQGRHHVQEPVFPFATR
SRRRSASASTPSSPASVDPMEEVTPKSVSTPSVDLSQDQEGMDTQQVDAESTY
SDTEYQDDKBFIGIDLVMGKIKFSPWAMVVMKATSKRQAMPKRWQWFGDGKF
SEIKLVALGLFQHOFLNATFNKLVSYRKAMVHTLEKARVAGTKTFSSPGESLED
QIKMLEWAGHGFPTGIEGLKPKQPVVVKSRSDSRNLEPRRRNKSRRTTN
DAASESPPKRLTNSVGGKDRGDEEERBMASVETNNKLEDRCLSCGKNPV
FHPLEGGCSCDRPFLFMYDEDCYQCTVCEGKELLCSNTSCRECEVC
LEVLVAGTADAKLQEPWCMCLPQRCCHVLRDRNRQDFITDPLDEEFP
PGQIPAIAPAKRKIRLVISFDGIATGVLVLELGIKVKRYIASEVCABEIAVGTGH
EKQIKVNDVRKRTIKKNIEEWGFDLVIGGSPCNDSNVNPKARKGYEGTGRFFEFY
HLNLYTRPKEDGNRPFPFMFENVAMKNDKKDISRLACPNVIMDAIKVSAARHRY
PWGNLPGMNRPMVMSKNDKLEQDCLFRTAKLKVQITITTKNSGIRQGNOLFVY
MNGKDDVLWCTELERIFGPFAHYTDVSNMNGRGRQKLLGRSVIPRIHLPAPLKDYF
ACE"
```

BASE COUNT 1051 a 1084 c 1147 g 941 t

ORIGIN

Alignment Scores:

Pred. No.: 1,926-248 Length: 4223  
Score: 4573.00 Matches: 857  
Percent Similarity: 99.77% Conservative: 0  
Best Local Similarity: 99.77% Mismatches: 2  
Query Match: 99.63% Indels: 0  
DB: 10 Gaps: 0

US-09-720-086-6 (1-859) x AFI51973 (1-4223)

Qy	1	MetLysGlyAspSerArgHisLeuAenGluGluGluGluAlaSerGlyTyrGluGluCys	20
Db	297	ATGAAGGGAGACAGACAGATCTGAATGAAGAAGAGGGTCCAGCGGGTATGAGGAGTGC	356
Qy	21	IleIleValAenGlyAenPheSerAspGlnSerSerAspThrLysAspAlaProSerPro	40
Db	357	ATTATCGTTAATGGGAACCTTCAGTGACCACTCTCAGACACGAGGATGCTCCCTCACCC	416
Qy	41	ProValLeuGluAlaIleCysThrGluProValCysThrProGluThrArgGlyArgArg	60
Db	417	CCAGTCTTGGAGGCAATCTGCACAGAGCCAGTCTGCACACAGAGACAGAGGCGGAGG	476
Qy	61	SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuAenTyrThrGlnAspMet	80
Db	477	TCAAGCTCCCGGCTGTCTAAGAGGGAGGTCTCCAGGCTTCTGAATTACACGAGACATG	536
Qy	81	ThrGlyAspGlyAspArgAspGluValAspAspGlyAenGlySerAspIleLeuMet	100
Db	537	ACAGAGATGGAGACAGAGATGATGAAGTAGATGATGGGAATGGCTCTGATATTCTAATG	596
Qy	101	ProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSerProAlaValArg	120
Db	597	CCAAAGCTCACCGGTGAGACCAAGGACACAGAGCGGCTCTGAAAGCCCGGCTGTCCGA	656
Qy	121	ThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnArgAlaSerProArgIleThr	140
Db	657	ACCCACATAGCAATGGGACCTCCAGCTTGGAGAGCAAGAGCCCTCCCGAGAAATCAC	716
Qy	141	ArgGlyArgGlnGlyArgHisValGlnGluTyrProValGluPheProAlaThrArg	160
Db	717	CGAGGTCCGCGGCGCCACCATGTGACAGGAGTACCTCTGGAGATTTCGGGCTTACCAGG	776
Qy	161	SerArgArgGlnArgAlaSerSerAlaSerThrProTrpSerSerProAlaSerVal	180
Db	777	TCTCGAGACGTGAGACATCGTCTTCAGCAGACAGCCCATGGTCATCCCTGCCAGCGTC	836
Qy	181	AspPheMetGluGluValThrProLysSerValSerThrProSerValAspLeuSerGln	200
Db	837	GACTTTCATGGAAGAGTGAACCTAAGAGCGTCAAGACCATACCCCATCAGTTGACTTGA	896

Qy	201	AspGlyAspGlnGluGlyMetAspThrThrGlnValAspAlaGluSerArgAspGlyAsp	220
Db	897	GATGAGATCAGAGAGGTATGGATACACACAGGTGATGCGAGAGCATATATGAGAC	956
Qy	221	SerThrGluTyrGlnAspAspLysGluPheGlyIleGlyAspLeuValTrpGlyLysIle	240
Db	957	AGCAGACAGATATCAGATGATAAAGATTTGGAAATAGGTGACCTCGTGTGGGAAAGATC	1016
Qy	241	LysGlyPheSerTrpTrpProAlaMetValValSerTrpLysAlaThrSerLysArgGln	260
Db	1017	AAGGGCTTCTCTCGTGGCTGCTCCATGGTGTCTCTGGAAAGCCACCTCCCAAGCGACAG	1076
Qy	261	AlaMetProGlyMetArgTrpValGlnTrpPheGlyAspGlyLysPheSerGluLysSer	280
Db	1077	GCATGCCCGGAATCGCTGGGTACAGTGGTGGTATGGAAGTTTCTGAGATCTCT	1136
Qy	281	AlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsn	300
Db	1137	GCTGACAACTGGTGGCTCTGGGGCTCTGAGCCAGCACCTTAAATCTGGCTTACCTTCAAT	1196
Qy	301	LysLeuValSerTyrArgLysAlaMetTyrHisThrLeuGluLysAlaArgValArgAla	320
Db	1197	AAGCTGGTTCTTATAGGAAGGCCATGTACACACTCTGGAGAAGCCAGGCTTCGAGCT	1256
Qy	321	GlyLysThrPheSerSerProGlyGluSerLeuGluAspGlnLeuLysProMetLeu	340
Db	1257	GGCAAGACCTTCTCCAGCAGTCTCTGGAGAGTCACTGGAGGACCACTGGAAGCCCATCTG	1316
Qy	341	GluTrpAlaHisGlyGlyPheLysProThrGlyIleGluGlyLeuLysProAsnLysLys	360
Db	1317	GAGTGGGCCACCGTGGCTTCAAGCTTCTGGATCGAGGCCCTCAAAACCAACAAAGAG	1376
Qy	361	GlnProValValAsnLysSerLysValArgArgSerAspSerArgAsnLeuGluProArg	380
Db	1377	CAACAGCTGTTAATAGTCGAAGGTGCGTCTTACAGACAGTAGGAACTTAGAACCCAGG	1436
Qy	381	ArgArgGluAsnLysSerArgArgArgThrThrAsnAspSerAlaAlaSerGluSerPro	400
Db	1437	AGACGCGAGAACAAAGTCGAGACGACCAACCAATGCTCTGCTTCTGAGTCCGCC	1496
Qy	401	ProProLysArgLeuLysThrAsnSerTyrGlyLysAspArgGlyGluAspGluGlu	420
Db	1497	CCACCCAGCGGCTCAAGACAAATAGCTATGGCGGGAAGACCCGAGGGAGGATGAGAG	1556
Qy	421	SerArgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAspArgCys	440
Db	1557	AGCCGAGAACGGATGGCTTCTGAAGTCACCAACAAAGGCAATCTGGAAGACCGCTGT	1616
Qy	441	LeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyLysCys	460
Db	1617	TTGTCTGTGGAAAGAAAGAACCTGTGTCTTCCACCCCTCTTTGAGGGTGGGCTCTGT	1676
Qy	461	GlnSerCysArgAspArgPheLeuGluLeuPheTyrMetTyrAspGluAspGlyTyrGln	480
Db	1677	CAGAGTTCGCGGATCGCTTCTTAGAGCTTCTTACATGTATGATGAGACGGCTATCAG	1736
Qy	481	SerTyrCysThrValCysCysGluGlyArgGluLeuLeuLeuCysSerAsnThrSerCys	500
Db	1737	TCCTACTGACCGCTGTGTGTGAGGCGGTGAACCTGTGTGTGTGAGTGAACCAAGTGC	1796
Qy	501	CysArgCysPheCysValGluCysLeuGluValLeuValGlyAlaGlyThrAlaGluAsp	520
Db	1797	TGCAGATGTTCTGTGTGGAGTGTCTGGAGTGTCTGGGCGCGAGGCACAGCTGAGAT	1856
Qy	521	AlaLysLeuGluGluProTrpSerCysTyrMetCysLeuProGlnArgCysHisGlyVal	540
Db	1857	GCCAGGTGACAGAACCTTGGAGCTGCTATATGTGCTTCCCTCAGCGCTGCATGGGTC	1916
Qy	541	LeuArgArgArgLysAspTrpAsnMetArgLeuGlnAspPheThrThrAspProAsp	560
Db	1917	CTCCAGCGCAGGAAAGATTGGAAATGCGGCTCGAAGACTTCTTCACTACTGATCCTGAC	1976

```

Oy 561 leuGluc1uPheGluProProluYrProAlaIleProAlaAlaIlyArGArgPro 580
Db 1977 CTGGAAGAAATTGAGCCACCCAGTTGTACCCGCAATTCCTTGACCCAAAAGAGAGGCC
Oy 581 lIearGValIleSerleuPheAspGlyIleAlaThrGlyTyrIleValIleuysGluLeu 600
Db 2037 ATTAGAGTCTCTCTCTCTGTTGATGGAATTGCACGGGGTACTTGCTGCTCAAGAGACTTG
Oy 601 GlyIleuysValGluIlyTyrIleAlaSerGluValCysAlaGluSerIleAlaValGly 620
Db 2097 GGATTTAAAGTGGAAAGTACATTGCTCCGAGAGTCTGTGCAAGACTCCATCGCTGTGGGA
Oy 621 ThrAllyuShieGluIlyGluIleIlyTyrValAsnAspValArgIleIleThrIlylys 640
Db 2157 ACTGTTAAGCATGAGAGCCAGATCAATATGTCATGACGTCGGAAATACCAAGAGAAA
Oy 641 AsnIleGluIlyTyrProluPheAspLeuValIleGlyGlySerProCysAsnAspLeu 660
Db 2217 AATATTGAAGAGTGGGGCCCTTCGACTTGATGATGTTGTTGGAAGCCCATGATGATCTC
Oy 661 SerAsnValAsnProAlaArgIlyGlyLeuTyrGluIlyThrGlyArgIlyPheGlu 680
Db 2277 TCTAACGTCATCTCTGCGCCGAAAGTTTATAGAGGGCACAGAAAGCTCTTCTTCGAG
Oy 681 PheTyrHisIleuLeuAsnTyrThrArgProGlyGluIlyAspAsnArgProPhePheTyr 700
Db 2337 TTTTACCACTTCTGTAATTAATACCCGCCCAAGAGAGGGCGACACCTTCATCTTCTGG
Oy 701 MetPheGluAsnValAlaIleMetIlyValAsnAspIlyAspIleSerArgPheLeu 720
Db 2397 ATGTTGGAATGTTTGGCCATGAAAGATGACAAAGAAAGACATCCAGATTCCTG
Oy 721 AlaCysAsnProValMetIleAspAlaIleIlyValSerAlaIleIlyArgAlaArgTyr 740
Db 2457 GCATGTAACCCAGTGTATGATGATGCCATCAAGCTGTCTGCTCACAGGGCCCGGTAC
Oy 741 PheTyrGlyAsnLeuProGlyMetAsnArgProValMetAlaSerIlyAsnAspIlyLeu 760
Db 2517 TTTCGGGGTAACCTTACCCGGAATGAACGGCCCTGTATGCTTCAAGATGATTAACCTC
Oy 761 GluLeuGluAsnProCysLeuGluIlyPheSerArgThrAlaIlyIleuIlyValGluIle 780
Db 2577 GAGCTGCAGGAGTCCCTGGAGTTCAGTGACAGCAAAAGTTAAAGAGTGCAGACATA
Oy 781 ThrThrIlySerAsnSerIleArgGluIlyIlyAsnGluLeuPheProValIleMetAsn 800
Db 2637 ACCACCAAGTCACTCCATCAGACAGGCAAAAACAGCTTTTCCCTGATGATGAAT
Oy 801 GlyIlyAspAspValIleuTyrPheThrGluLeuGluIlyIlePheGlyPheProAlaHis 820
Db 2697 GGCAAGAGAGCAGCTTTTGTGCTGCACTGAGCTCGAAAGGATCTTGGCTTCCCTGCTCAC
Oy 821 TyrThrAspValSerAsnMetGlyArgGlyAlaArgGluIlyLeuLeuGlyArgSerTyr 840
Db 2757 TACACGAGAGCTGTCACATAGGCGCGCGCGCTCAGAAAGCTGCTGGCAGAGTCTCGG
Oy 841 SerValProValIleArgHisLeuPheAlaProLeuIlyAspTyrPheAlaCysGlu 859
Db 2817 AGTGAACGGGTATCAGACACCTGTTGCCCTTGAAGAGACTTCTTGGCTGTGAA

```

```

AUTHORS Yin, B., Chen, Y.T., Zhu, M., Luo, Y.J., Zhu, N., Xu, S.C., Wu, G.Y. and
Shen, Y.
TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing
isoforms in mouse embryonic tissue
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4338)
AUTHORS Yin, B., Chen, Y.T., Zhu, M., Luo, Y.J., Zhu, N., Xu, S.C., Wu, G.Y. and
Shen, Y.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular
Biology, Institution of Basic Medical Sciences, Chinese Academy of
Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,
P.R. China
FEATURES
source
1..4338
/organism="Mus musculus"
/strain="KM"
/db_xref="taxon:10090"
/dev_stage="8-9 day old embryo"
1..4338
/gene="Dnmt3b"
412..2991
/gene="Dnmt3b"
/function="methylates cytosine in DNA"
/note="alternatively spliced"
/codon_start=1
/product="DNA cytosine-specific methyltransferase isoform
1"
/protein_id="AAF74515.1"
/db_xref="GI:8347118"
/translation="MGDSRLNREEGASGYECITVNGFNFSDDSDTADAPSPVLE
ALTEPVCTPERTGRSSRLSKREVSSILANTYQDMTGDGDRDDEVDGNGSDILMPK
LTREKDKTRTESPAVTRHNSGTSLEKRPASPRITRGQGRHVOEYVPEPATR
SRRRASSDSSTPWSBPASVDPMEBVTPEKSPSPVLSODGDGMDTQVDAESIY
GDSREYQDKEKRGIDGLVWGKITGFSWPMAMVSWKTSKROAMGMRVOWFEDGKF
SEISADKVALGLFSGHFNATFNKLVSYRKMYHTLEKARVAKTFSSPGESLED
QLKPLLMWAGGFKPTGIEGLKPNKQPVNKSQVRSDSNLEPRRREKSRRTTN
DSASESPPKRLKTNISYGGKRGDEBEREASVTNNKGNLEDRCSGKNPVAS
FHPLEGGICOSGRDRELLEYMDDEGYOYCTVCGEGRLLCSNTGCRGCEVC
LEVLVAGTAEADKLLQEPWSCMCLPORCHVLRKDKMNRLODFTTDPDLEPEP
PKLYPAIPAKRRPIRLSLFDGATGLVLEKGLIVERKIASVCAESIAGTVKA
EGQIKYVNDVRKITTKNIEWGPFDLVIGSPNDLSVNPARKGLYEGGRLEFEFY
HLINLYTPKEGDNRPPEWMPENVMKVKDKDISRLACNPVMDAIVKSAHRARY
FWGNLPGNMRFVWASKNDKLELDQCLBERSRTAKLKYQVITTTKSNISROCKNQLPVV
MNGKDDVLMCTELERLFGFPAHYTDVSNNGRARKLGRSWSVPVIRHLFAPLDYF
ACE"
BASE COUNT 1073 a 1124 c 1173 g 968 t
ORIGIN
Alignment Scores:
Pred. No.: 1,986-248 Length: 4338
Score: 4573.00 Matches: 857
Percent Similarity: 99.77% Conservative: 0
Best Local Similarity: 99.77% Mismatches: 2
Query Match: 99.63% Indels: 0
Gaps: 0
US-09-720-086-6 (1-859) x AFI51969 (1-4338)
Oy 1 MetIySGIyAspSerArgHisIleuAsnGluIlyGluIlyAlaSerGlyTyrGluIlyCys 20
Db 412 ATGAAGGAGAGACAGACATCTGAATGAAGAAGAGGCTCCACGGGTATAGAGAGTGC 471
Oy 21 lIelleValAsnGlyIlyAsnPheSerAspGluIlySerSerAspThrIlyAspAlaProSerPro 40
Db 472 ATTAATCGTTAATGGGAATTCAGTGACCAAGTCTCTAGACACGAAGAGATGCTCCCTCAACC 531
Oy 41 ProValIleuGluAlaIleCysThrGluIlyProValCysThrProGluThrArgGlyArgArg 60
Db 532 CCAGCTTTGGAGGCAATCTGCACAGAGCAAGTCTGCACACCAAGACCAAGAGCCGCGAGG 591
Oy 61 SerSerArgLeuSerIlyArgGluIlyAlaSerSerLeuIlyAsnTyrThrGluIlyMet 80

```

Db 592 TCAAGCTCCCGGCTGTCTTAAGAGGAGGTCTCCAGCCTTCTGAATTACACGCGAGCATG 651  
Qy 81 ThrGlyAspGlyAspArgAspGluValAspAspGlyAsnGlySerAspIleLeuMet 100  
Db 652 ACAGGAGATGAGACAGAGATGATCAAGTAGATGATGGAAATGGCTCTGATATTTCTAATG 711  
Qy 101 ProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSerProAlaValArg 120  
Db 712 CCAAGGCTCACCGGTGAGCAAGGACACAGGACCGCTCTGAAAGCCCGGCTGCCGA 771  
Qy 121 ThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnArgAlaSerProAlaIleThr 140  
Db 772 ACCCGACATAGCAATGGGACCTCCAGCTTGGAGAGGCAAGAGCCCTCCCGCAGATCAC 831  
Qy 141 ArgGlyArgGlnGlyArgHisValGlnGluTyrProValGluPheProAlaThrArg 160  
Db 832 CGAGGTCCGACGGGCGCCACCATGTGACGAGTACCTGTGGAGTTTCCGGCTACACAG 891  
Qy 161 SerArgArgArgAlaSerSerAlaSerThrProTrpSerSerProAlaSerVal 180  
Db 892 TCTCGAGACGTCGAGCATCGTCTTCAGCAAGCACCCCATGGTTCATCCCTGCCAGCGTC 951  
Qy 181 AspPheMetGluGluValThrProLysSerValSerThrProSerValAspLeuSerGln 200  
Db 952 GACTTCATGGAGAAGTGCACCTAAGACGGTCAGTACCCCATCAGTTGACTTGAGCCAG 1011  
Qy 201 AspGlyAspGlnGlyMetAspThrThrGlnValAspAlaGluSerArgAspGlyAsp 220  
Db 1012 GATGGAGATCAGGAGGATATGGATATCCACAGAGTGGATGCAGAGCATATATGGAGAC 1071  
Qy 221 SerThrGluTyrGlnAspAspLysGluPheGlyIleGlyAspLeuValTrpGlyLysIle 240  
Db 1072 AGCACAGAGTATCAGGATGATTAAGAGTTTGGAAATAGGTACCTCGTGTGGGAAAGATC 1131  
Qy 241 LysGlyPheSerTrpTrpProAlaMetValSerTrpLysAlaThrSerLysArgGln 260  
Db 1132 AAGGGCTTCTCTGGTGGCTCCCATGGTGGTGTCTCTGGAAAGCCACCTCAAGCCGACAG 1191  
Qy 261 AlaMetProGlyMetArgTrpValGlnTrpPheGlyAspGlyLysPheSerGluIleSer 280  
Db 1192 GCCATGCCCGGAATGCGTGGGTACAGTGGTTGGTGTATGGCAAGTTTCTCGAGATCTCT 1251  
Qy 281 AlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsn 300  
Db 1252 GCTGACAAACTGGTGGCTCTGGGCTGTTCAGCCAGCACCTTTAATCTGGCTACCTTCAAT 1311  
Qy 301 LysLeuValSerTyrArgLysAlaMetTyrHisThrLeuGluLysAlaArgValArgAla 320  
Db 1312 AAGCTGGTTCTTATAGGAAGGCCATGTACCACTCTGGAGAAAGCCAGGGTTTCGAGCT 1371  
Qy 321 GlyLysThrPheSerSerProGlyGluSerLeuGluAspGlnLeuLysProMetLeu 340  
Db 1372 GCAAGACCTTCTCCAGCAGTCTCTGAGAGTCACTGGAGGCCAGCTGAAGCCCATGCTG 1431  
Qy 341 GluTrpAlaHisGlyGlyPheLysProThrGlyIleGluGlyLeuLysProAsnLysLys 360  
Db 1432 GAGTGGGCCCCGCTGGCTTCAAGCCTACTGGGATCGAGGGCCCTCAAAACCCAAAGAAAG 1491  
Qy 361 GlnProValValAsnLysSerLysValArgArgSerAspSerArgAsnLeuGluProArg 380  
Db 1492 CAACCAAGTGTAAATAGTCGAAGTGGTGGTTCAGACAGTAGGAACTTAGAACCCACAG 1551  
Qy 381 ArgArgGluAsnLysSerArgArgArgThrThrAsnAspSerAlaIleSerGluSerPro 400  
Db 1552 AGACCGGAGACAAAGTCAAGACCGCAACCAATGACTCTGCTGCTTCTGAGTCCCCC 1611  
Qy 401 ProProLysArgLeuLysThrAsnSerTyrGlyLysAspArgGlyGluAspGluGlu 420  
Db 1612 CCACCAAGCGCTCAAGACAAATAGCTATGGCGGGAAGGACCCGAGGGAGATGAGGAG 1671  
Qy 421 SerArgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAspArgCys 440  
Db 1672 ACCCGGAGAACGGATGGCTTCTGAAGTCACTCAACCAAGGGCCATCTGGAGAGACCGCTGT 1731

Qy 441 LeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyGlyLeuCys 460  
Db 1732 TTGTCTCTGTGAAAGAAAGACCTGTGTCTTCCACCCCTCTTTGAGGGTGGCTCTGT 1791  
Qy 461 GlnSerCysArgAspArgPheLeuGluLeuPheTyrMetTyrAspGluAspGlyTyrGln 480  
Db 1792 CAGAGTTGCCGGATCGCTTCTAGAGCTTCTCATCATGTATGATGAGGACGGCTATCAG 1851  
Qy 481 SerTyrCysThrValCysCysGluGlyArgGluLeuLeuLeuCysSerAsnThrSerCys 500  
Db 1852 TCTTACTACACCGTGTGTGTGAGGGCCGTGAACGTCTGTCTGTGACGTAAACACAGGTGC 1911  
Qy 501 CysArgCysPheCysValGluCysLeuGluValLeuValGlyAlaGlyThrAlaGluAsp 520  
Db 1912 TGCAGATCTTCTGTGTGAGTGTCTGGAGGTGTCTGGTGGCGGACGACAGCTGAGGAT 1971  
Qy 521 AlaLysLeuGlnGluProTrpSerCysTyrMetCysLeuProGlnArgCysHisGlyVal 540  
Db 1972 GCCAAGCTGCAGGAACCTCTGGAGCTGTATATGTGCTCTCCCTCAGCGCTGCCATGGGTC 2031  
Qy 541 LeuArgArgArgLysAspTrpAsnMetArgLeuGlnAspPhePheThrThrAspProAsp 560  
Db 2032 CTCCGACGACGAGAAAGATTTGGAACATGCCCTGCAAGACTTCTTCACTGATCTCTGAC 2091  
Qy 561 LeuGluGluPheGluProProLysLeuTyrProAlaIleProAlaLysArgArgPro 580  
Db 2092 CTGGAAGAAATTTGAGCCACCCCAAGTTGTATCCAGCAATTTCTCTCAGCCCAAGAGGCC 2151  
Qy 581 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeu 600  
Db 2152 ATTAGAGCTCTGTCTCTGTGTGATGGAATGCAACGGGGTACTTGGTGTCTCAAGAGTTG 2211  
Qy 601 GlyIleLysValGluLysTyrIleAlaSerGluValCysAlaGluSerIleAlaValGly 620  
Db 2212 GGTATTAAGTGAAGTAAAGTACATTTGCTCCGAAGTCTGTGCAGAGTCCATCTGTGGGA 2271  
Qy 621 ThrValLysHisGluGlyGlnIleLysTyrValAsnAspValArgLysIleThrLysLys 640  
Db 2272 ACTGTTAAGCATGAAGGCCAGATCAATATGTCAATGACGTCCGGAANAATCACCAGNAA 2331  
Qy 641 AsnIleGluGluTrpGlyProPheAspLeuValIleGlyGlySerProCysAsnAspLeu 660  
Db 2332 AATATTGAAGAGTGGGGCCGTTTCGACTTGTGATTTGGTGAAGCCCATGCAATGATCTC 2391  
Qy 661 SerAsnValAsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPhePheGlu 680  
Db 2392 TCTAAGCTCAATCTCTCCCGCAAGGTTTATATGAGGGCACAGAGAGCTCTTCTTCGAG 2451  
Qy 681 PheTyrHisLeuLeuAsnTyrThrArgProLysGluGlyAspAsnArgProPhePheTrp 700  
Db 2452 TTTTACCACCTTGTGAATTTATACCCCGCCCAAGGGCGGACACCCCTCCATTTCTTCGG 2511  
Qy 701 MetPheGluAsnValValAlaMetLysValAsnAspLysLysAspIleSerArgPheLeu 720  
Db 2512 ATGTTGAGAAATGTTGTGGCCATGAAAGTGAATGACAAAGAAAGACATCTCAAGATTCTG 2571  
Qy 721 AlaCysAsnProValMetIleAspAlaIleLysValSerAlaAlaHisArgAlaArgTyr 740  
Db 2572 GCATGTAAACCCAGTATGATGATGATGATCAAGGTGTCTGTCTCAGGGCCCGGTAC 2631  
Qy 741 PheTrpGlyAsnLeuProGlyMetAsnArgProValMetAlaSerLysAsnAspLysLeu 760  
Db 2632 TTCTGGGGTAACTCACTCCCGGAATGAACAGGCCCTGTGTGCTTCAAGAAATGATAAGCTC 2691  
Qy 761 GluLeuGluAspCysLeuGluPheSerArgThrAlaLysLeuLysValGlnThrIle 780  
Db 2692 GAGCTCAGAGCTGCTCTGGAGTTCAAGTAGGACAGCAAGGTTAAAGAAAGTGCAGACAATA 2751  
Qy 781 ThrThrLysSerAsnSerIleArgGlnGlyLysAsnGlnLeuPheProValValMetAsn 800  
Db 2752 ACCACCAAGTCAAGTCCATCAGACAGGGCAAAACCAAGCTTTTCCCTGTAGTCATGAAT 2811



Qy 801 G1YlYsAsPAsPValLeuTTPCySerThrgluLeuGluuArgilePheGlyPheProAlaHis 820  
Db 2812 GGCAAGAGCAGAGCTTTTGTGTGCTGAGCTGGAAGAGATCTTGCGCTTCCCTGCTCAC 2871  
Qy 821 TTTThraPvalSerAsnMeG1YArgG1YAlaArgG1YnYsLeuLeuG1YArgSerTirp 840  
Db 2872 TACCGGAGCTGTCCACACATGGGCGCGCGCGCTGACAGAGCTGCTGGGCGAGCTCTGG 2931  
Qy 841 SerValProValIleArgHisLeuPheAlaProLeuYsAspTyrPheAlaCysGlu 859  
Db 2932 AGTGTACCGGTATCATGACACCTGTTTGGCCCTTGAAGAGCTTACTTGTGCTGTGAA 2988  
RESULT 4  
AF068627 4135 bp mRNA linear ROD 06-DEC-1999  
LOCUS AF068627  
DEFINITION Mus musculus DNA cytosine-5 methyltransferase 3B2 (Dnmt3b) mRNA,  
alternatively spliced, complete cds.  
ACCESSION AF068627  
VERSION AF068627.2 GI:6449471  
KEYWORDS  
SOURCE Mus musculus.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS Okano, M., Xie, S. and Li, E.  
TITLE Cloning and characterization of a family of novel mammalian DNA  
(cytosine-5) methyltransferases  
Nat. Genet. 19 (3), 219-220 (1998)  
JOURNAL 98324766  
MEDLINE 9662389  
PUBMED  
REFERENCE  
AUTHORS Xie, S., Okano, M. and Li, E.  
TITLE Direct Submision  
JOURNAL Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,  
Charlestown, MA 02129, USA  
3 (bases 1 to 4135)  
Okano, M., Chijiwa, T., Sasaki, H. and Li, E.  
Direct Submision  
JOURNAL Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,  
Charlestown, MA 02129, USA  
REMARK  
COMMENT Sequence update by submitter  
FEATURES  
On Nov 18, 1999 this sequence version replaced gi:3327981.  
Location/Qualifiers  
1. .4135  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/map="84.0 cM"  
1. .4135  
/gene="Dnmt3b"  
1. .4135  
/gene="Dnmt3b"  
/note="similar to EST sequences deposited in GenBank  
Accession Numbers AA116694, AA119979, AA177277, AA210568,  
AA407106, and AA515617"  
269. .2788  
/gene="Dnmt3b"  
/function="de novo DNA methylation"  
/note="alternatively spliced product; contains Cys-rich  
region; C-terminal region is similar to corresponding  
region of DNA cytosine-5 methyltransferase 3A"  
/codon\_start=1  
/product="DNA cytosine-5 methyltransferase 3B2"  
/protein\_id="AAC40179.2"  
/db\_xref="GI:6449472"  
/translation="MKGDSRLHNEEGAGSEECIYVNGNFGDSSTKADSPVLE  
AICTEPTCPETRRGRSSRLSKREVSLINTQDMGDDGDDDEVDDNGSGDILMPK  
LTRERDTRRSBPAYATRHNGTSLERORASPRITRGRGRHVQETPVERPATR  
SRRRASSASTPWSVPASVDPMEVTPKSVSTPSVDLSQDQDEMDTTQVDAESRD  
GSDTEYODDEKFGIGDLYWGKIKGFSWMPAMVSWKATSKROAMPKRWOMFGDGK  
SEISADKLVALGLFSOHPLATFNKLVSRYKAMVHTLSEKARVAGTFFSSPGLSD  
QLKPMLEMAHGGFKPTGIGLKPKNKGPKNSRRRTTNDASASESPPKLTKTNSYGG

KDRGDESERREMASEVTNNKLNLEDRLCSGKKNPVSPHPLFEEGLCOSCRPFLFL  
FYHIDEDGOSYCTVCEBGEFLILCSNNSCCRFCEVCELEVLVGAGTAEDAKIOEPS  
CYMCLPQCHGVARRDMNMRLODFETDPLEEBEPKLYPA1PAKRPIRLVL  
FDGIATGLYLVKELEGIKEVXIASEVCAESIAGTVHBEQIKYVNDVKITKNIIE  
WGPDLVIGGSPCNDISNVNPARKGLYSTGRLFEEYHLINTYTPKEGDRPEFMF  
ENVYAMKRNKDIDSRFLACNPVMDALKVSAAHRAPIPKMNPGRMNRPMASKNDKL  
ELQDCLERSRTAKLVQYTTTKNSINQKQOLFVYVMNGKDVLCWTELEIIFGFP  
AHYTDVSNMGRGAROKLGRSMSPVIRHLFAPLKDYFACE"  
BASE COUNT 1043 a 1048 c 1106 g 938 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 9,09e-243 Length: 4135  
Score: 4472.00 Matches: 839  
Percent Similarity: 97.67% Conservative: 0  
Best Local Similarity: 97.67% Mismatches: 0  
Query Match: 97.43% Indels: 20  
DB: 10 Gaps: 1  
US-09-720-086-6 (1-859) x AF068627 (1-4135)  
Qy 1 MetlyGc1YAsPAsPValLeuTTPCySerThrgluLeuGluuArgilePheProAlaHis 20  
Db 269 ATGAAGGAGACAGCAGACATCTGAATGAAGAGGAGGTCACAGCGGTATGAGAGTGC 328  
Qy 21 11e11eValaAng1YAsnPheSerAspGlnSerSerAspThryAspAlaProSerPro 40  
Db 329 ATTATCGTTAATGGGAACCTTCAGTACCAAGTCTTCAGACAGAGAGATGCTCCACCC 388  
Qy 41 ProValLeuGluuAla11eCysThrGluProValCysThrProGluThrArgG1YArg 60  
Db 389 CCAAGCTTGAGGCAATCTGACAGAGCAGCTGACACAGAGAGCCAGAGCCGACAG 448  
Qy 61 SerSerSerArgLeuSerArgG1YValSerSerLeuLeuAnTYrThrgluAspMet 80  
Db 449 TCAAGCTCCCGCTGTCTAAGAGGAGGCTTCAGCTTCTGAATTACAGCAGACATG 508  
Qy 81 ThrG1YAsPAsPValLeuTTPCySerThrgluLeuGluuArgilePheProAlaHis 100  
Db 509 ACAAGAGATGAGACAGAGATGAGATGATGATGATGATGATGATGATGATGATGATG 568  
Qy 101 ProlyLeuThrArgG1YThryAsPAsPValLeuTTPCySerThrgluLeuGluuArg 120  
Db 569 CCAAGCTCACCGTGTAGACCAAGAGCAGAGAGCGCTTGAAGCCGCGCTGTCCGA 628  
Qy 121 ThrArgHisSerArgG1YThrySerSerLeuGluuArgG1YArgAlaSerProArgIleThr 140  
Db 629 ACCGACATAGCAATGGAGCTCCAGCTTGAGAGGCAAGAGCTCCCGCAATCACC 688  
Qy 141 ArgG1YArgG1YArgHisIshIsvaG1YArgIYrProValGluPheProAlaThrArg 160  
Db 689 CGAGGTGCGCAGGCGCCGCAACATGTGCAAGAGTACCTGTGAGATTCCCGCTACCA 748  
Qy 161 SerArgArgArgArgAlaSerSerSerAlaSerThrProTyrSerSerProAlaSerVal 180  
Db 749 TCTCGAGACGTCAGACATGCTCTTCAACAGACAGCCATGTATATCCCTCCACGCTC 808  
Qy 181 AspPheMetGluGluValThrProlySerValSerThrProSerValAspLeuSerGln 200  
Db 809 GACTTCATGAGAGAGATGACACCTTAAGAGCTCAGTACCCCATATGATGATGAGAC 868  
Qy 201 AspG1YAsPAsPValLeuTTPCySerThrgluLeuGluuArgilePheProAlaHis 220  
Db 869 GATGAGATCAGAGAGGTATGATACCAACAGGTGATGATGATGATGATGATGATGATG 928  
Qy 221 SerThrgluYrGluAspAspLySgluPheGlyIleG1YAspLeuValTTPG1YLyS11e 240  
Db 929 AGCACAAGATATCAGATGATTAAGAAGATTGGAATAGTGACCTCGTGTGGGAAAGATC 988  
Qy 241 LyseG1YPheSerTirpProAlaMetVala1SerTirpLySAlaThrSerLySArgGln 260  
Db 989 AAGGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1048



QY	261	AlaMetProGlyMetArgTrpValGlnTrpPheGlyAspGlyLysPheSerGluIleSer	280
Db	1049	GCCATGCCCGAATCGCGTGGTACAGTGGTTGGTGGTATGTCGCAAGTTTCTCGAGTCTCT	1108
QY	281	AlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsn	300
Db	1109	GCTGACAAACTGGTGGCTCTGGGGCTGTTCAGCCAGCACTTTAATCTGGCTACTCTCAAT	1168
QY	301	LysLeuValSerTyrArgLysAlaMetTyrHisThrLeuGluLysAlaArgValArgAla	320
Db	1169	AAGCTGGTTCTTATAGGAAGGCCATGTACCACTCTCGAGAAAGCCAGGGTTCGAGCT	1228
QY	321	GlyLysThrPheSerSerProGlyGluSerLeuGluAspGlnLeuLysProMetLeu	340
Db	1229	GGCAAGACCTTCTCAGAGTCTCGAGAGTCACTGGAGACCACTGAAGCCCATGCTG	1288
QY	341	GluTrpAlaHisGlyGlyPheLysProThrGlyIleGluGlyLeuLysProAsnLysLys	360
Db	1289	GAGTGGGCCCCACGGTGGCTTCAAGCCTACTGGGATCGAGGGCTCAAAACCAAGAAAG	1348
QY	361	GlnProValValAsnLysSerLysValArgSerAspSerArgAsnLeuGluProArg	380
Db	1349	CAACCA-----	1354
QY	381	ArgArgGluAsnLysSerArgArgThrThrAsnAspSerAlaLysSerGluSerPro	400
Db	1355	-----GAGAACAAAGTCAAGACGACACCAACCAATGACTCTGCTTCTGAGTCCCCC	1408
QY	401	ProProLysArgLeuLysThrAsnSerTyrGlyLysAspArgGlyGluAspGluGlu	420
Db	1409	CCACCAAGCGCTCAAGACAAATAGCTATGGCGGAAGACCCAGCGGAGGATGAGGAG	1468
QY	421	SerArgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAspArgCys	440
Db	1469	AGCCAGAGACCGATGGCTTCTGAAGTCAACCAACCAAGGCAATCTGGAAGACCGCTGT	1528
QY	441	LeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyGlyLeuCys	460
Db	1529	TGTCTCTGGGAAAGAAACCCCTGTCTTCCACCCCTCTTTGAGGGTGGCTCTGT	1588
QY	461	GlnSerCysArgAspArgPheLeuGluLeuPheTyrMetTyrAspGluAspGlyTyrGln	480
Db	1589	CAGAGTTCGGGATCGCTTCTTAGAGCTTCTACATGATGATGAGGACGGCTATCAG	1648
QY	481	SerTyrCysThrValCysCysGlyArgGluLeuLeuLeuLeuCysSerAsnThrSerCys	500
Db	1649	TCCTACTGCACCGTGTCTGTGAGGCGCGTGAATGCTGTCTGTGCAAGTAAACCAAGCTGC	1708
QY	501	CysArgCysPheCysValGluCysLeuGluValLeuValGlyAlaGlyThrAlaGluAsp	520
Db	1709	TGCAGATGTTCTGTGTGAGTGTCTGAGGTGCTGGTGGCGCAGGCACAGCTGAGGAT	1768
QY	521	AlaLysLeuGlnGluProThrSerCysTyrMetCysLeuProGlnArgCysHisGlyVal	540
Db	1769	GCCAAAGTGCAGAAACCCCTGGAGCTGCTATATGTGCTTCCCTCAGCGCTGCATGGGTC	1828
QY	541	LeuArgArgArgLysAspTrpAsnMetArgLeuGlnAspPhePheThrAspProAsp	560
Db	1829	CTCCGACGCGAGAAAGATTGGAACATCGCGCTGCAAGACTTCTTCACTACTGATCTGAC	1888
QY	561	LeuGluGluPheGluProProLysLeuTyrProAlaThrLeuPheAlaLysArgArgPro	580
Db	1889	CTGGAAGAATTGAGCCACCAAGTTGTACCAAGCAATCTTCAGCCAAAGAGGCC	1948
QY	581	IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeu	600
Db	1949	ATTAGAGTCTCTCTCTGTTTGTGTAATGGAATTCGAACGGGGTACTTGTGCTCAAGGAGTGT	2008
QY	601	GlyIleLysValGluLysTyrIleAlaSerGluValCysAlaGluSerIleAlaValGly	620
Db	2009	GGTATTAAAGTGGNAAGTACATTGCCCTCCGAAGTCTGTGCAGAGCTCCATCGCTGGGA	2068
QY	621	ThrValLysHisGluGlyGlnIleLysTyrValAsnAspValArgLysIleThrLysLys	640

Db	2069	ACTGTTAAGCATGAGGCCAGATCAATATATGTCAATGACGTCCGGAAATCACCAGAA	2128
QY	641	AsnIleGluGluTrpGlyProPheAspLeuValIleGlyGlySerProCysAsnAspLeu	660
Db	2129	AATATTGAAGAGTGGGCCCGTTCGACTTGGTGAITGGTGAAGCCCATGCAATGATCTC	2188
QY	661	SerAsnValAsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPheGlu	680
Db	2189	TCTAACGTCATCTTCCCGCCGCAAGGTTTATATAGGGGCACAGGAAGGCTCTTCTTCGAG	2248
QY	681	PheTyrHisLeuLeuAsnTyrThrArgProLysGluGlyAspAsnArgProPhePheTrp	700
Db	2249	TTTTACCACTTGTCTGAATATATACCCGCCCAAGAGGGGACAAACCGTCCATTTCTTGG	2308
QY	701	MetPheGluAsnValAlaMetLysValAsnAspLysLysAspLysArgPheLeu	720
Db	2309	ATGTTTCGAGAAATGTTGTGGCCATGAAAGTGAATGACAAAGAACACATCTCAAGATTCTG	2368
QY	721	AlaCysAsnProValMetIleAspAlaIleLysValSerAlaAlaHisArgAlaArgTyr	740
Db	2369	GCATGTAAACCCAGTGTATGTCATGCAAGGTGTCTGCTCTCACAGGGCCCCGGTAC	2428
QY	741	PheTrpGlyAsnLeuProGlyMetAsnArgProValMetAlaSerLysAsnAspLysLeu	760
Db	2429	TTCGGGGTAACTTACCCGGAATGAACAGGCCCGTGTATGGCTTCAAGAAATGATAAGCTC	2488
QY	761	GluLeuGlnAspCysLeuGluPheSerArgThrAlaLysLeuLysValGlnThrIle	780
Db	2489	GAGCTGCAGGACTGCTGGAGTTTCAGTAGACAGCAAAAGTTAAAGAAAGTGCAGACAATA	2548
QY	781	ThrThrLysSerAsnSerIleArgGlnGlyLysAsnGlnLeuPheProValValMetAsn	800
Db	2549	ACCACCACTCGAACTCCATCAGACAGGGCAAAACCCAGCTTTTCCCTGTAGTATGAAT	2608
QY	801	GlyLysAspAspValLeuTrpCysThrGluLeuGluArgIlePheGlyPheProAlaHis	820
Db	2609	GGCAAGGACAGCTTTTGTGGTCACTGAGCTCGAAGGATCTTCGGCTTCCCTGCTCAC	2668
QY	821	TyrThrAspValSerAsnMetGlyArgGlyAlaArgGlnLysLeuLeuGlyArgSerTrp	840
Db	2669	TACACGAGCTGTCCAAACATGGGCCCGCGCTCAGAGCTGTGGGAGGTCCTCTGG	2728
QY	841	SerValProValIleArgHisLeuPheAlaProLeuLysAspTyrPheAlaCysGlu	859
Db	2729	AGTGACCGGTCTATCAGACACCTGTTTGGCCCTTGAAGGACTACTTTCCTGTGAA	2785

RESULT 5			
AF151974	4163 bp	mRNA	linear
LOCUS	Mus musculus DNA cytosine-specific methyltransferase isoform 6		
DEFINITION	(Dnmt3b) mRNA, complete cds.		
ACCESSION	AF151974		
VERSION	AF151974.1	GI:8347130	
KEYWORDS			
SOURCE	Mus musculus.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	1. (bases 1 to 4163)		
	Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and		
	Shen, Y.		
TITLE	Cloning of full-length Dnmt3b cDNA and its alternative splicing		
	isoforms in mouse embryonic tissue		
JOURNAL	Unpublished		
REFERENCE	2. (bases 1 to 4163)		
AUTHORS	Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and		
	Shen, Y.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-MAY-1999); Department of Biochemistry and Molecular		
	Biology, Institution of Basic Medical Sciences, Chinese Academy of		
	Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,		
	P.R.China		

FEATURES  
source  
Location/Qualifiers  
1..4163  
/organism="Mus musculus"  
/strain="KM"  
/db\_xref="taxon:10090"  
/dev\_stage="8-9 day old embryo"  
1..4163  
/gene="Dmrt3b"  
297..2816  
/gene="Dmrt3b"  
/function="methylates cytosine in DNA"  
/note="alternatively spliced"  
/codon\_start=1  
/product="DNA cytosine-specific methyltransferase isoform 6"  
/protein\_id="AA074520.1"  
/db\_xref="GI:8347131"  
/translation="MKGDSRLNEERAGSGYEELIYVNGNPSDOSDTRKDAAPPVLE  
ALCTEPCVCTPETERGRSSRLSKREYSSLLNTYODMGDGRDDEVDGGSDLMPK  
LIRETKDTIRSSPAVTRHSNGTSSLERQASPRIIRGRGRHHVOEYPIVEPATR  
SRRRASASSATPMSVDFMEVTPKSVSPVLSODGDEMDTQVDESIY  
GDSYQDDKEFGIGDLVMGKIKGFSWPMAMVWSKATSKQAMPQMRWQWFGDGK  
SEISADKVALGILPSQHFNLAFTNCLVSTRKAMHTLEKARACGTSSSPESLED  
QLEPMLBWAHGFRRPTGIBGLPNKKQEPENKSRRTYNDASASEPPEKRLKTNISYGG  
KDRGDEBEREMASVETNNKGNLEDRCLSCGKNPSPFHPLEGLCOSCRDFLEL  
PYWDEDEYOSYCTVCCGCEGRELILCSNTSCRCPCVEGLEVLVAGTAENAKIQEPM  
CYMCLPQRCHGVARRKDMNRLOPFTTDPLEPEPEPKLIPALPAARPRIPRLVLSL  
FDGATGVLVLEKGIKEVKYIASVECAESIAGTVHNEGIKYVNDVRKITKNIEE  
WGPPDLVIGSPNDLSNPNPARKLYEGTGRLFFEFYHLNLTYPKEGNRPFPMMF  
ENNVAMLVNDKDISRELACNPVMDIKYSAHARARFWGMLPGNRPMAASNDKL  
ELQCLFESRTAKLKVKQITTTKSNISROGKNQLFPVYMNCGDVLMCELERLFGFP  
AHYTDVSNNGRGARQKULGRMSVPIVIRHLFAPLKDYFACE"

BASE COUNT 1032 a 1074 c 1128 g 929 t

ORIGIN

Alignment Scores:  
Pred. No.: 8.3e-242 Length: 4163  
Score: 4455.00 Matches: 837  
Percent Similarity: 97.44% Conservative: 0  
Best Local Similarity: 97.44% Mismatches: 2  
Query Match: 97.06% Indels: 20  
DB: 10 Gaps: 1

US-09-720-086-6 (1-859) x AF151974 (1-4163)

Q 1 MetLysGlyAspSerArgHisLeuAsnGluGluGluValAspGlyTyrGluGluCys 20  
Db 297 ATGAAGGAGACAGACGACATCTGAATAGAGAGAGGTCACCGGCTATGAGAGAGTGC 356  
Qy 21 ILeIleValAsnGlyAsnPheserAspGlnSerSerAspThrLysAspAlaProSerPro 40  
Db 357 ATTATCGTTAATGGGAATCTCAGTACCAAGTCTTCAGACACAGAGAGATGCTCCCTCAACC 416  
Qy 41 ProValLeuGluAlaIleCysThrGluProValCysThrProGluThrArgGlyArgArg 60  
Db 417 CCAGCTTGGAGGCAATCTGCACAGAGCCAGTCTGCACACACAGAGACCAAGAGCCGAGG 476  
Qy 61 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuAsnTyrThrGlnAspMet 80  
Db 477 TCAAGCTCCCGCTGTCTAAGAGGAGAGGCTCTCAGCTTCTGAATTAACGCGAGGACATG 536  
Qy 81 ThrGlyAspGlyAspArgAspAspGluValAspAspGlyAsnGlySerAspIleLeuMet 100  
Db 537 ACAAGAGATGAGACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 596  
Qy 101 ProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSerProAlaValArg 120  
Db 597 CCAAGGCTACCGGTAGACCAAGACACCAAGACCCGCTCTGAAAGCCGCGCTGTCCGA 656  
Qy 121 ThrArgHisSerAsnGlyThrSerSerLeuGluArgGluArgAlaSerProArgIleThr 140  
Db 657 ACCGCACTATGACATGAGACCTCAGCTTGGAGAGGCAAGAGGCTCCCGCAGATACCC 716

Qy 141 ArgGlyArgGluGluValArgHisIleValGluGluTyrProValGluPheProAlaThrArg 160  
Db 717 CGAGGTGCGCAGGCGCGCCACCAAGTGCAGAGTACCTGTGGATTTCGGGTACACAGG 776  
Qy 161 SerArgArgArgArgAlaSerSerSerAlaSerThrProThrAspSerProAlaSerVal 180  
Db 777 TCTCGAGACGTCGACATCTGCTTCAAGACACGCGCAAGTGTATCCCTCGCAGGCTC 836  
Qy 181 AspPheMetGluGluValThrProLysSerValSerThrProSerValAspLeuSerGln 200  
Db 837 GACTTCATGAAAGAAATGACACCTTAAGAGCGTCAGTACCCCATAGTACCTTGAGCCAG 896  
Qy 201 AspGlyAspGluGluGluValMetAspThrThrArgGluValAspAlaGluSerArgAspGlyAsp 220  
Db 897 GATGGAGATCAGAGAGGATGATGATACCAACAGTGTATCAGAGAGATATATGGAGAC 956  
Qy 221 SerThrGluTyrGluAspAspLysGluPheGlyIleGlyAspLeuValTyrGlyLysIle 240  
Db 957 AGCACAGATATCAGATATGATTAAGATTTGGAATAGTACCTGTGTGGGAAAGATC 1016  
Qy 241 LysGlyPheSerTyrProAlaMetValSerTyrLysAlaThrSerLysArgGln 260  
Db 1017 AAGGCTTCTCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1076  
Qy 261 AlMetProGluMetArgTyrPvalGlnTyrPheGlyAspGlyLysPheSerGluIleSer 280  
Db 1077 GCCATGCCCGGATGCGCTGGGATCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1136  
Qy 281 AlAspLysLeuValAlaLeuGlyLysPheSerGlnHisPheAsnLeuAlaThrPheAsn 300  
Db 1137 GCTGACAAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1196  
Qy 301 LysLeuValSerTyrArgLysAlaMetTyrHisThrLeuGluLysAlaArgValArgAla 320  
Db 1197 AAGCTGCTTCTTATGAGAGGCGCATGTACACACTCTGAGAGAAAGCGGTTCCAGCT 1256  
Qy 321 GlyLysThrPheSerSerSerProGlyLysSerLeuGluAspGlnLeuLysProMetLeu 340  
Db 1257 GCGAAGACCTTCTCCAGCAGTCTCTGAGAGTACTGAGAGACCGAGCTAAGCCCATCTG 1316  
Qy 341 GluTyrAlaHisGlyGlyPheLysProThrGlyIleGluGlyLysLeuLysProAsnLys 360  
Db 1317 GAGTGGGCGCCACGCTGCTTCAAGCCTTACTGGGATGAGGAGGCTTCAAAACCAAGAG 1376  
Qy 361 GlnProValValAsnLysSerLysValArgArgSerAspSerArgAsnLeuGluProArg 380  
Db 1377 CAACCC----- 1382  
Qy 381 ArgArgLysAsnLysSerArgArgArgThrThrAsnAspSerAlaAlaSerGluSerPro 400  
Db 1383 -----GAGAACAAAGTGCAGAGCGCACCAACCATGACTCTGCTTCTGAGTCCCCC 1436  
Qy 401 ProProLysArgLeuLysThrAsnSerTyrGlyGlyLysAspArgGlyGluAspGluGlu 420  
Db 1437 CCACCCAAAGCGCTCAAGACAAATATGCTATGCGGAGAGACCAAGGAGAGATGAGAG 1496  
Qy 421 SerArgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAspArgCys 440  
Db 1497 AGCCGAGAACGATGCTTCTGAAGTCAACCAACCAAGGCAATCTGGAAGACCGCTGT 1556  
Qy 441 LeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyLysLeuCys 460  
Db 1557 TTGCTCTGTGAAAGAAAGAACCGTGTCTTCCACCCCTCTTGAAGGTGGCTCTGT 1616  
Qy 461 GlnSerCysArgAspArgPheLeuGluLeuPheTyrMetTyrAspGluAspGlyTyrGln 480  
Db 1617 CAGAGTTGCCCGGATCGCTTCTTGAAGTCTTCAATGATATGATGAGAGACGGCTATCAG 1676  
Qy 481 SerTyrCysThrValCysCysGluGluValArgGluLeuLeuLeuCysSerAsnThrSerCys 500  
Db 1677 TCTACTGACACCGCTGCTGTGAGGCGCGTGAACGTGCTGTGACAGTAAACCAAGCTGC 1736  
Qy 501 CysArgCysPheCysValGluCysLeuGluValLeuValGluValGluValThrAlaGluAsp 520

LOCUS AF151970 4278 bp mRNA linear ROD 08-JUN-2000  
DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 2  
(Dnmt3b) mRNA, complete cds.  
ACCESSION AF151970  
VERSION AF151970.1 GI:83471119  
KEYWORDS  
SOURCE Mus musculus.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 4278)  
AUTHORS Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and  
Shen,Y.  
TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing  
isoforms in mouse embryonic tissue  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 4278)  
AUTHORS Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and  
Shen,Y.  
TITLE Direct Submission  
JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular  
Biology, Institution of Basic Medical Sciences, Chinese Academy of  
Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,  
P.R.China  
FEATURES  
Location/Qualifiers  
1..4278  
/organism="Mus musculus"  
/strain="KM"  
/db\_xref="taxon:10090"  
/dev\_stage="8-9 day old embryo"  
1..4278  
/gene="Dnmt3b"  
412..2931  
/gene="Dnmt3b"  
/function="methylates cytosine in DNA"  
/note="alternatively spliced"  
/codon\_start=1  
/product="DNA cytosine-specific methyltransferase isoform  
2"  
/protein\_id="AAF74516.1"  
/db\_xref="GI:83471120"  
/translation="MKGDSRRHLNEEGASGYEECIIVNGNFSDOSSDTKDAPSPVLE  
ALCTPVCPTETRRSSRLSKREVSLLNYTQDMTGDDGDDDDVDGNGSDILMPK  
LRTREKDTTRTSAPVTRSHNGTSLERQASPRITRQGRHGVQYVPEFPA  
SPRRASSASPTWSPASVDMEVTPKSVSTPSVDLSQDQDGMDDTQVDAES  
GSTEVQDDKEFGIGDLVNGKIKGFSWPMVMSWKATSKQAMPGRVQMGDK  
SEISADKLVALGLFSQHPNLATFNKLVSRKAMVHTLEKARVAGTKTSSPESLE  
QKPMLEWAHGGFPKPTGIGLKNKKQPKENRRRTTNDSSAESPPLKRLKTNSTGG  
KDRGDEESEREMASEVTNNKGLNEDRLCSGKNPVSFPLFEGGLCQSCDRLEL  
FYMYDEGYQSYCTVCEGRELCCNTSCRCFCVECLVLAGAGTAEDAKLQEPWS  
CYMCPQRCGLVLRKRNRLQDFFTPDLEFPKPLYPALPAKRRIRVLSL  
FDGIATGLVLKELGIVKEKYIASVCAESIAGTVKHEGQIKYVNDVRKTKKQIEE  
WGFPLVIGGSPCNLSNVNPAKGLVEGTGRLPFFVYHLLNLTYPKGNRPFFMMP  
ENVVAKVNDKKDISRFLACNPVMDAIKVSAAHARYFWGNLPGMNPVMAKNDKL  
ELQDCLFSPRTAKLKKVQITTKNSITROGKNOLFVVMNGKDDVLWCTELERIFGFP  
AHYTVSNMGRGARQKLLGSRMSVYPVIRHLFAPLKDYPACE"

BASE COUNT 1055 a 1113 c 1154 g 956 t  
ORIGIN

Alignment Scores:  
Pred. No.: 8,58e-242 Length: 4278  
Score: 4455.00 Matches: 837  
Percent Similarity: 97.44% Conservative: 0  
Best Local Similarity: 97.44% Mismatches: 2  
Query Match: 97.06% Indels: 20  
Db: 10 Gaps: 1

US-09-720-086-6 (1-859) x AF151970 (1-4278)

QY 1 MetLysGlyAspSerArgHisLeuAsnGluGluGluGlyAlaSerGlyTyrGluGluCys 20  
DB 412 ATGAAGGGAGACAGACATCTGAATGAAGAGAGGGTGCAGCGGTATGAGGAGTGC 471-

QY 21 IleIleValAsnGlyAsnPhseSerAspGlnSerSerAspThrLysAspAlaProSerPro 40  
Db 472 ATTATCGTAAATGGGAACCTTCAGTACCACTCTCAACACAGAGATCTCCCTCAACC 531  
QY 41 ProValIleuGluAlaIleCysThrGluProValCysThrProGluThrArgGlyArg 60  
Db 532 CCACTCTGGAGGCAATCTGCACAGACCACTGTGCACACAGAGACCGAGGCGCAGG 591  
QY 61 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuAsnTyrThrGlnAspMet 80  
Db 592 TCAAGCTCCGGCTGTCTAAGAGGAGGTCTCCAGCTTCTGAATTACCGCAGGACATG 651  
QY 81 ThrGlyAspGlyAspArgAspAspGluValAspAspGlyAsnGlySerAspIleLeuMet 100  
Db 652 ACAGAGAGTGGAGACAGAGATGATAGATGATGATGATGATGATGATGATGATGATG 711  
QY 101 ProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSerProAlaValArg 120  
Db 712 CCAAGCTCACCGGTGAGACCAAGGACCAAGGACCGGCTGTGAAGCCCGGCTGTCCGA 771  
QY 121 ThrArgHisSerAsnGlyThrSerSerLeuGluArgGluArgAlaSerProArgIleThr 140  
Db 772 ACCGACATAGCAATGGACCTCCAGCTTGGAGAGGAAAGCCTCCCGCAGATCAC 831  
QY 141 ArgGlyArgGluGlyArgHisIleValGlnGlyTyrProValGluPheProAlaThrArg 160  
Db 832 CGAGGTGGGAGGCGCGCCACCATGTGCAGAGTACCTGTGGAGTTTCCGGCTACAG 891  
QY 161 SerArgArgArgArgAlaSerSerSerAlaSerThrProTyrSerSerProAlaSerVal 180  
Db 892 TCTCGAAGACGTGAGACATGCTTTCAGCAGACGACGCGATGCTATCCCTGCAGGCTC 951  
QY 181 AspPheMetGluGluValThrProLysSerValSerThrProSerValAspLeuSerGln 200  
Db 952 GACTTCATGGAAGAAGTGAACACCTTAAGAGCGTCACTACCCATCGTTGACTTGAGCCAG 1011  
QY 201 AspGlyAspGluGluGlyMetAspThrThrGlnValAspAlaGluSerArgAspGlyAsp 220  
Db 1012 GATGAGATCAGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 1071  
QY 221 SerThrGluTyrGlnAspAspLysGluPheGlyIleGlyAspLeuValTyrGlyLysIle 240  
Db 1072 AGCACAGAGTATCAGAGTATGATTAAGATTGGAATAGTACCTGTGTGGGAAAGATC 1131  
QY 241 LysGlyPheSerTyrTyrProAlaMetValValSerTyrLysAlaThrSerLysArgGln 260  
Db 1132 AAGGCTCTCTCGTGGGCTGCGCATGTGTGTCTGCGAAGCCACCTCCAAAGCAG 1191  
QY 261 AlaMetProGlyMetArgTyrTyrValGlnTyrPheGlyAspGlyLysPheSerGluIleSer 280  
Db 1192 GCCATGCCCGAATGCGCTGGGTACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1251  
QY 281 AlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsn 300  
Db 1252 GCTGACAAACTGT 1311  
QY 301 LysLeuValSerTyrArgLysAlaMetTyrHisThrLeuGluLysValAlaArgValArgAla 320  
Db 1312 AAGCTGT 1371  
QY 321 GlyLysThrPheSerSerSerProGlyGluSerLeuGluAspGluLeuLysProMetLeu 340  
Db 1372 GGCAGAGCTTCTCCAGCAGCTCTGAGAGTCACTGAGAGACCACTGAAAGCCATGCTG 1431  
QY 341 GlyThrAlaHisGlyGlyPheLysProThrGlyTyrLeuGlyLysLeuLysProAsnLysLys 360  
Db 1432 GAGTGGGCGCCAGGTGTGTCAAGGCTTACTGTGGATCGAGGCGCTCAAAACCAAGAG 1491  
QY 361 GlnProValValAsnLysSerLysValArgArgSerAspSerArgAsnLeuGluProArg 380  
Db 1492 CAACCA----- 1497  
QY 381 ArgArgGluAsnLysSerArgArgArgThrThrAsnAspSerAlaAlaSerGluSerPro 400

Db 1498 -----GAGAACAAAGTCCAGAGCGCACACCAATGACTTGTGCTTGTGAGTCCCC 1551  
QY 401 ProProLysArgLeuLysThrAsnSerTyrGlyLysAspArgGlyValAspGluGlu 420  
Db 1552 CCACCCAAAGCGCTCAAGACAAATAGCTATGGCGGAGAGACCCAGGAGATGAGAGAG 1611  
QY 421 SerArgIleArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAspArgCys 440  
Db 1612 AGCGGAGAACGATGCTTGTGAAGTCAACCAACAAAGGCAATCTGGAAGCCGCTGT 1671  
QY 441 LeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyLysCys 460  
Db 1672 TTGTCTCTGGAAGAAGAACCTGTGTCTTCCACCCCTCTTGAAGGTGGCTGTGT 1731  
QY 461 GlnSerCysArgAspArgPheLeuGluLeuPheTyrMetTyrAspGluAspGlyTyrGln 480  
Db 1732 CAGAGTCCGGGATGCTTCTTGAAGCTTCTTACATGATGATGAGAGCGGCTATCAG 1791  
QY 481 SerTyrCysThrValCysCysGluGlyArgGluLeuLeuLeuCysSerAsnThrSerCys 500  
Db 1792 TCTACTGCACCGGT 1851  
QY 501 CysArgCysPheCysValGluCysLeuGluValLeuValGlyAlaGlyThrAlaGluAsp 520  
Db 1852 TGCAATGCTTCTGT 1911  
QY 521 AlaLysLeuGluGluProTyrSerCysTyrMetCysLeuProGluArgCysHisGlyVal 540  
Db 1912 GCCAAGCTGCAGGAACCCCTGAGCTGATATATGTGCTCTCCAGCGCTGCCATGGGCTC 1971  
QY 541 LeuArgArgArgLysAspTyrPheMetArgLeuGlnAspPheThrThrAspProAsp 560  
Db 1972 CTCGAGCGCAGGAAGAAATGGAACATGCGCTGCAGAGACTTCTTCACTGATCTGAC 2031  
QY 561 LeuGluGluPheGluProProLysLeuTyrProAlaIleProAlaAlaLysArgArgPro 580  
Db 2032 CTGGAAGAAATTTGAGCACCAACCAATGTATGCCAGCAATCTCTGAGCCAAAGAGGCC 2091  
QY 581 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeu 600  
Db 2092 ATTAGAGTCTGTCTGT 2151  
QY 601 GlyIleLysValGluLysTyrIleAlaSerGluValCysAlaGluSerIleAlaValGly 620  
Db 2152 GGTATTTAAGTGAAGAAATGATGCTCCGAAAGTGTGTGAGAGTCCATCGCTGTGGGA 2211  
QY 621 ThrValLysHisGluGlyGlnIleLysTyrValAsnAspValArgLysIleThrLysLys 640  
Db 2212 ACTGTTAAGCATGAAGGCGAGATCAAAATATGTCAATGACCGCCGAAATCAACCAAAA 2271  
QY 641 AsnIleGluGluTyrGlyProPheAspLeuValIleGlyLysSerProCysAsnAspLeu 660  
Db 2272 AATATTGAAGAGTGGGCGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2331  
QY 661 SerAsnValAsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPhePheGlu 680  
Db 2332 TCTAAGCTCAATCTCTCCCGCAAGGTTATATGAGGGCAGAGAGGCTCTTCTTTCGAG 2391  
QY 681 PheTyrHisLeuLeuAsnTyrThrArgProLysGluGlyAspAsnArgProPhePheTyr 700  
Db 2392 TTTTACCACTTGTGTAATTTATACCGCCCAAGAGGCGCAACCGTCAATCTTGTGG 2451  
QY 701 MetPheGluAsnValValAlaMetLysValAsnAspLysLysAspIleSerArgPheLeu 720  
Db 2452 ATGTTGAGAAATGTTGTGGCATGAAGTGAAGACAAAGAAACATTCAAATTCCTG 2511  
QY 721 AlaCysAsnProValMetIleAspAlaIleLysValSerIleAlaHisArgAlaArgTyr 740  
Db 2512 GCATGTAAACCGAGTGAATGATGAGCCATCAAGGTGTGTGTGTGTGTGTGTGTGTGTGT 2571  
QY 741 PheTyrGlyAsnLeuProGlyMetAsnArgProValMetAlaSerLysAsnAspLysLeu 760

Db 2572 TTCTGGGGTAACCTACCCGGAATGAACAGCCCGGTGATGGCTTCAAGAATGATAAGCTC 2631  
QY 761 GluLeuGlnAspCysLeuGluPheSerArgThrAlaLysLeuLysLysValGlnThrIle 780  
Db 2632 GAGCTGCAGGACTGCTGGAGTTTCAGTAGCAGACAAAGTTAAAGAAAGTGCAGACAATA 2691  
QY 781 ThrThrLysSerAsnSerIleArgGlnGlyLysAsnGlnLeuPheProValValMetAsn 800  
Db 2692 ACCACCAAGTCGAATCCATCAGACAGGCAAAACAGCTTTTCCCTGTAGTCATGAAT 2751  
QY 801 GlyLysAspValLeuTrpCysThrGluLeuGluArgIlePheGlyPheProAlaHis 820  
Db 2752 GCGAAGGACGAGCTTTTGTGGTGCACCTGAGCTCGCAAGGATCTTCGGCTTCCCTGCTCAC 2811  
QY 821 TyrThrAspValSerAsnMetGlyArgGlyAlaArgGlnLysLeuLysLeuGlyArgSerTrp 840  
Db 2812 TACAGGACGTGTCCACATGGCCGCGCGCCCGTCAGAAAGCTGCTGGGAGGCTCCTGG 2871  
QY 841 SerValProValIleArgHisLeuPheAlaProLeuLysAspTyrPheAlaCysGlu 859  
Db 2872 AGTGATCCCGGTTCATCAGACACCTGTTTGGCCCTTGAAGGACTACTTTTGCCTGTGAA 2928  
RESULT 7  
AY078427 4006 bp mRNA linear ROD 01-JUL-2002  
LOCUS AY078427 Mus musculus DNA cytosine methyltransferase 3b6 (Dnmt3b6) mRNA, complete cds; alternatively spliced.  
DEFINITION  
ACCESSION AY078427  
VERSION AY078427.1 GI:21655120  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus  
REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
TITLE Dnmt3a2, encoded by transcripts from an intronic promoter of Dnmt3a is expressed abundantly in ES cells and germ cells  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 4006)  
AUTHORS Chen, T., Ueda, Y. and Li, E.  
TITLE Direct Submission  
SUBMITTED (08-FEB-2002) Cardiovascular Research Center,  
MASSACHUSETTS General Hospital, 149 13th Street, Charlestown, MA 02129, USA  
FEATURES  
source  
1. 4006  
/organism="Mus musculus"  
/strain="129S4/SvJae"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/map="84.0 CM"  
1. 4006  
/gene="Dnmt3b6"  
269..2659  
/gene="Dnmt3b6"  
/function="de novo methylation"  
/note="contains PHD and PWWP domains; alternatively spliced"  
/codon\_start=1  
/product="DNA cytosine methyltransferase 3b6"  
/protein\_id="AA185481.1"  
/db\_xref="GI:21655121"  
/translation="MKGSGRHLNEEGASGYEECIIVNGNFSQSDTKDAPSPVLE  
LIREPTVCTPGRSSRLSKREVSLNLTQDMTGDGRDDEVDVNGSDILMPK  
ATREKTRTESPAVTRHNSGSSLRQASPRITRQGRHHVQEPYVFPATR  
SRRRASSTPWSVSPASVDFMEETPKSVSTVDLSQDQEGHMDTTQVDAESRD  
GDSTEYQDDKEFGIDLVGKIKFSPWPMVYSWKATSKRQAMPGRWQWFDGKF  
SEISADKLVALGFLFSQHENLATFNKLVSYRKAMVHTLEKARVRAGTKTSSPSLES  
QLKPMLEWAGGFKPTGIEGLNKKQPVVYKSKVRSSDRLNLEPRRKNKSRRTTN  
DSASPPPKRLKNSYGGKDRGEDBSRERMASEVTNNKGNLEDCSLSCGKNPV  
FHLPEGLCGSCDRDFLEFLFYMDYDQYQSVCTVCEGRELILLNNTSCRCFCV  
LEVLVAGTAGDAKLQEPWSCVNCPLPQCHGVLRKRDWMLQDFFTTDPDLEEFEP

PKLYPAIPAAKRRIPIVLSLFDGIATGYLVKELGKVEKYIASVCAESIAVGTVKH  
EQIYKVNDRKITTKNIIEWGPFDLVIGGSPCNLDNPNVARKGLYGTGTLFFEFY  
HLLNTRPKEGNRPFPMFENVAMKNDKIDISRLACNPVMTDAIKVSAHRARY  
FWGNLPGMNRIFGPPAHYTDVSNNGRAGROKLLGRSWSPVIRHLFAPLKDYFACE"  
BASE COUNT 1001 a 1017 c 1075 g 913 t  
ORIGIN  
Alignment Scores: 8.56e-229 Length: 4006  
Pred. No.: 4223.50 Matches: 796  
Score: 4223.50  
Percent Similarity: 92.67% Conservative: 0  
Best Local Similarity: 92.67% Mismatches: 0  
Query Match: 92.02% Indels: 63  
DB: 10 Gaps: 1  
US-09-720-086-6 (1-859) x AY078427 (1-4006)  
QY 1 MetLysGlyAspSerArgHisLeuAsnGluGluGluGlyTyrGluGluCys 20  
Db 269 ATGAAGGGAGACAGCAGACATCTGAATGAAGAAGGGTGCAGCGGTATGAGGAGTGC 328  
QY 21 IleIleValAsnGlyAsnPheSerAspGlnSerSerAspThrLysAspAlaProSerPro 40  
Db 329 ATTATCGTTAATGGAACTTCAGTGACCACTCTCAGACACAGAGGATGCTCCCTCACCC 388  
QY 41 ProValLeuGluAlaIleCysThrGluProValCysThrProGluThrArgGlyArgArg 60  
Db 389 CCAGTCTTGAGGCAATCTGCACAGACCCAGTCTGCACACAGAGACAGAGCCGCGCAGG 448  
QY 61 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuAsnTyrThrGlnAspMet 80  
Db 449 TCAAGCTCCCGGTGCTAAGAGGGAGGTCTCCAGCCTTCTGAATTACACGAGACATG 508  
QY 81 ThrGlyAspGlyAspAspAspGluValAspAspGlyAsnGlySerAspIleLeuMet 100  
Db 509 ACAGGAGATGGAGACAGAGATGATGAAGTAGATGAGGATGGGATGGCTCTGATATTCTAATG 568  
QY 101 ProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSerProAlaValArg 120  
Db 569 CCAAGCTCACCGGTGAGACCAAGGACCAAGGACCGGCTCTGAAAGCGCGGCTGTCCGA 628  
QY 121 ThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnArgAlaSerProArgIleThr 140  
Db 629 ACCGACATAGCANTGGGACCTCCAGCTTGGAGAGGCAAGAGCCTCCCCAGATCACC 688  
QY 141 ArgGlyArgGlnGlyArgHisValGlnGluTyrProValGluPheProAlaThrArg 160  
Db 689 CGAGGTCCGCGAGGCGCCACCATGTGACAGAGTACCCCTGTGGAGTTCCTCCGCTACCCAG 748  
QY 161 SerArgArgArgAlaSerSerSerAlaSerThrProTrpSerSerProAlaSerVal 180  
Db 749 TCTCGAGAGCTCGAGCATCGTCTTCAGCAAGCAGCCATGGTCATCCCTCCAGCGTC 808  
QY 181 AspPheMetGluGluValThrProLysSerValSerThrProSerValAspLeuSerGln 200  
Db 809 GACTTCATGGAAGAAGTACACCTTAAGAGCGTCACTACCCCATCAGTTGACTTGAGCCAG 868  
QY 201 AspGlyAspGlnGluGlyMetAspThrThrGlnValAspAlaGluSerArgAspGlyAsp 220  
Db 869 GATGGAGATCAGGAGGTATGGATACCAACAGCTGGATGATGAGAGCAGAGATGGAGAC 928  
QY 221 SerThrGluTyrGlnAspAspLysGluPheGlyIleGlyAspLeuValTrpGlyLysIle 240  
Db 929 ACCACAGATATCAGGATGATAAAGAGTTTGGAAATAGTAGTACCTCGTGTGGGAAAGATC 988  
QY 241 LysGlyPheSerTrpTrpProAlaMetValSerTrpLysAlaThrSerLysArgGln 260  
Db 989 AAGGGCTTCTCTCTGGTGGCCTGCATGGTGGTGTCTCTGAAAGCCACCTCCCAAGGACAG 1048  
QY 261 AlaMetProGlyMetArgTyrValGlnTrpPheGlyAspGlyLysPheSerGluIleSer 280  
Db 1049 GCCATGCCCGGAATGGCTGGGTACAGTGGTTGGTGTATGCAAGTCTTCTGAGATCTCT 1108

QY 281 AlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsn 300  
DB 1109 GCTACAAACTGGTGGCTGGGGCTGTCAGCCAGCATTTAACTGGCTACCTTCAT 1168  
QY 301 LysLeuValSerTyrArgLysAlaMetTyrHisThrLeuGlnLysValAlaArgValArgAla 320  
DB 1169 AAGCTGGTTCTTATAGGAAGGCCATGTACACACTCTGGAGAAAGCCAGGGTTCGAGCT 1228  
QY 321 GlyLysThrPheSerSerSerProGlyLysLeuGlnLysAspGlnLeuLysProMetLeu 340  
DB 1229 GGCAGACCTTCTCCAGCAGCTCGAGAGTCACTGGAGAGCCAGCTGAGCCAGCTG 1288  
QY 341 GlnTyrAlaHisGlyGlyPheLysProThrGlyLysLeuGlnLysProAsnLysLys 360  
DB 1289 GAGTGGCCCGACGGTGGCTTCAAGCTTACTGGGATCGAGGCCCTCAAAACCAAGAAAG 1348  
QY 361 GlnProValAlaAsnLysSerLysValArgArgSerAspSerArgAsnLeuGlnProArg 380  
DB 1349 CAACCAAGTGTATATAGTGAAGGTGCTGCTTCAGACAGTGAACCTTAGAACCCAGG 1408  
QY 381 ArgArgGlnAsnLysSerArgArgArgThrThrAsnAspSerAlaAlaSerGlnSerPro 400  
DB 1409 AGACGCGGAACAAAGTCCGAGACGACACACCAATGACTTGTCTTCTGAGTCCCC 1468  
QY 401 ProProLysArgLeuLysThrAsnSerTyrGlyGlyLysAspArgGlyValAspGlnGlu 420  
DB 1469 CCACCCAAAGCCCTCAAGACAAATAGCTTAGCGGGAGAGCCGAGGGAGATGAGAG 1528  
QY 421 SerArgGlnArgMetAlaSerGlnValThrAsnAsnLysGlyAsnLeuGlnAspArgCys 440  
DB 1529 AGCCGAGAGAGATGGCTTCTGAAGTCAACCAACAAAGGCATCTGGAGAGCCGCTGT 1588  
QY 441 LeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGlnGlyLysLeuCys 460  
DB 1589 TTGCTCTGTGAAGAAGAACCTGTGTCTTCCACCCCTCTTTGAGGTGGGCTGTGT 1648  
QY 461 GlnSerCysArgAspArgPheLeuGlnLeuPheTyrMetTyrAspGlnAspGlyTyrGln 480  
DB 1649 CAGAGTTCGGGGATGGCTTCTTACAGCTTCTTACATGATGAGAGCCGCTATCAG 1708  
QY 481 SerTyrCysThrValCysCysGlnGlyArgGlnLeuLeuLeuCysSerAsnThrSerCys 500  
DB 1709 TCCCTACTGACCGGTGTGTGTGAGGGCCGTGAAGCTGTGTGTGCACTGAACACAGCTGC 1768  
QY 501 CysArgCysPheCysValGlnCysLeuGlnValLeuValGlyAlaGlyThrAlaGlnAsp 520  
DB 1769 TGCAGATCTTCTGTGTGAGTGTGTGAGGTCTGTGGCGGACGACAGCTGAGAT 1828  
QY 521 AlaLysLeuGlnGlnProTyrPheSerCysTyrMetCysLeuProGlnArgCysHisGlyVal 540  
DB 1829 GCCAAGCTGCAGGAACCTGTGAGCTGTATATGTGCTCCCTCAGCGCTGCCATGGGGTC 1888  
QY 541 LeuArgArgArgLysAspTyrPheAsnMetArgLeuGlnAspPhePheThrThrAspProAsp 560  
DB 1889 CTCGAGCCAGGAAGATTGGAACATGGCCCTGCACAGACTTCTTCACTACTGATCCGAC 1948  
QY 561 LeuGlnGlnPheGlnProProLysLeuTyrProAlaIleProAlaAlaLysArgArgPro 580  
DB 1949 CTGGAAAGATTGAGCCACCAAGTTGATCCAGCATTCCTGCGAGCCAAAGAGAGGCC 2008  
QY 581 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyTyrLeuValLeuLysGlnLeu 600  
DB 2009 ATTAGAGTCTCTGTCTGTGATGGAATTGCAACGGGGTACTTGTGCTCAAGGAGTGTG 2068  
QY 601 GlyIleLysValGlnLysTyrIleAlaSerGlnValCysAlaGlnSerIleAlaValGly 620  
DB 2069 GGATTTAAAGTGAAGAAAGTACATGCTCCGAAGTCTGTGAGAGTCCATCGCTGTGGGA 2128  
QY 621 ThrValLysHisGlnGlnGlnIleLysTyrValAsnAspValArgLysIleThrLysLys 640  
DB 2129 ACTGTTAAGCATGAGGCGCAGATCAATATGTCAATGACGTCCGGAATAATCACCAAGAAA 2188  
QY 641 AsnIleGlnGlnIleTyrGlyProPheAspLeuValIleGlyLysSerProCysAsnAspLeu 660

DB 2189 AATATTGAAGAGTGGGGCCCGTTCGACTTGATTTGTGGAAACCCATGCAATGATCTC 2248  
QY 661 SerAsnValAsnProAlaArgLysGlyLeuTyrGlnGlyThrGlyArgLeuPheGln 680  
DB 2249 TCTAACGTCAATCTCCCGCAAGGTTTATATGAGGGCACAGGAGGCTCTTCTTCGAG 2308  
QY 681 PheTyrHisLeuLeuAsnTyrThrArgProLysGlnGlyAspAsnArgProPhePheTyr 700  
DB 2309 TTTTACACTTGTGTGATTTATATACCCGCCAAGAGGGCACAACCGTTCATTTCTGTG 2368  
QY 701 MetPheGlnAsnValAlaMetLysValAsnAspLysLysAspIleSerArgPheLeu 720  
DB 2369 ATGTTGGAATGTTGTGGCCATGAAGTGAATGACAGAAAGCATCTCAAGATTCTCTG 2428  
QY 721 AlaCysAsnProValMetIleAspAlaIleLysValSerAlaAlaHisArgAlaArgTyr 740  
DB 2429 GCATGTAAACCCAGTATGATGATGCCATCAAGGTGTCTGTCTGTACAGGGCCCGGTAC 2488  
QY 741 PheTyrGlyAsnLeuProGlyMetAsnArgProValMetAlaSerLysAsnAspLysLeu 760  
DB 2489 TTCTGGGTAACTTACCCGGAATGAC----- 2515  
QY 761 GlnLeuGlnAspCysLeuGlnPheSerArgThrAlaLysLeuLysValGlnThrIle 780  
DB 2515 ----- 2515  
QY 781 ThrThrLysSerAsnSerIleArgGlnGlyLysAsnGlnLeuPheProValAlaMetAsn 800  
DB 2515 ----- 2515  
QY 801 GlyLysAspAspValLeuTyrCysThrGlnLeuGlnArgIlePheGlyPheProAlaHis 820  
DB 2516 -----AGGATCTTCGGGCTTCCCTGTCTCAC 2539  
QY 821 TyrThrAspValSerAsnMetGlyArgGlyAlaArgGlnLysLeuLeuGlnYArgSerTyr 840  
DB 2540 TACACGAGCGTCTCCAACTGGGCGCGCGCCGCTCAGAGCTGTGGCAGGTCTGTGG 2599  
QY 841 SerValProValIleArgHisLeuPheAlaProLeuLysAspTyrPheAlaCysGln 859  
DB 2600 AGGTACCGGATCATGACACCTGTTGCCCTTGAAGACTACTTGGCTGTGAA 2656  
RESULT 8  
AF151975 4034 bp mRNA linear ROD 08-JUN-2000  
LOCUS AF151975  
DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 7  
(Dnmt3b) mRNA, complete cds.  
ACCESSION AF151975  
VERSION AF151975  
KEYWORDS AF151975.1 GI:8347134  
SOURCE Mus musculus.  
ORGANISM Mus musculus.  
REFERENCE 1 (bases 1 to 4034)  
AUTHORS Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and Shen,Y.  
TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing isoforms in mouse embryonic tissue  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 4034)  
AUTHORS Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and Shen,Y.  
TITLE Direct Submission  
JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular Biology, Institution of Basic Medical Sciences, Chinese Academy of Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005, P.R.China  
FEATURES  
source  
1..4034  
/organism="Mus musculus"  
/strain="KM"





```

Db 1857 GCCAAGCTGACGAGAACCCCTGGAGCTGTAATATGTGCTCCCTCAGCGCTGCATGGGGTC 1916
Qy 541 LeuAARGAARGLYSASPTRPASMMeLARGLeuGlnAspPhePheThrThraspProAsp 560
Db 1917 CTCGACCCAGAGAAAGATGGAACATGCGCCCTGCAGAGACTTCTTCACATCTGATCTCTGAC 1976
Qy 561 LeuGlnGluPheGluProProlyLeuTYRProAla1LeProAla1Ala1ysAgaArgPro 580
Db 1977 CTGAGAGAAATTGAGGCCACCCCAAGTTGTAACCCAGCAATTCCTGCAGCCAAAGAGAGCCC 2036
Qy 581 IleArgValIleSerLeuPheAspGlyIleAlaThrGlyTYRLeuValLeuLysGluLeu 600
Db 2037 ATTAGAGTCTGTCTGTTGTAATGGAATTGCAGCGGATCTGCTGCTCAAGAGATTG 2096
Qy 601 GlyIleLysValGluLysTYR1IleAlaSerGluValCysAlaGluSer1IleAlaValGly 620
Db 2097 GGATTTAAAGTGAAAAGTGCATTCCTCCGAAAGTCTGTGAGAGTCCATCCCTGTGGGA 2156
Qy 621 ThrValLysHisGluGluGlnIleLysTYRValAsnAspValArgLysIleThrLysLys 640
Db 2157 ACTGTTAAGCATGAGAGGCCAGATCAAAATATGTCATGACGTCGCGAAAATCACCAAGAAA 2216
Qy 641 Asn1IleGluGluTYRProPheAspLeuVal1IleGlyLysSerProCysAsnAspLeu 660
Db 2217 AATATTGAAGATGGGCGCCGTTGCACTTGATGATGGTGAAGGCCATGCAATGATCTC 2276
Qy 661 SerLeuValAsnProAlaArgLysGlyLeuTYRGLuGlyTYRGLYArgLeuPhePheGlu 680
Db 2277 TCTAACGCAATCTCCCGCAAGAGTTATATGAGGGGCAAGAGAGGCTCTCTTCTTGAG 2336
Qy 681 PheTYRHisLeuLeuAsnTYRThrArgProLysGluLysAspAsnArgProPhePheThr 700
Db 2337 TTTTACCACTTCTGTAATTATACCGCCGCAAGAGAGGCGCAACCGTCCATCTCTTG 2396
Qy 701 MetPheGluAsnValValAlaMetLysValAsnAspLysLysAsp1IleSerArgPheLeu 720
Db 2397 ATGTTGCGAAGATGTTGTGGCATGAAGTAAGCAAGAAAGACATCTCAAGATTCTG 2456
Qy 721 AlaCysAsnProValMetC1IleAspAla1IleLysValSerAla1Ala1HisArgAlaArgTYR 740
Db 2457 GCATGTAACCCAGTGAATGATCGATCGCCATCAAGGTGTCTGCTCACAGGCGCGGTAC 2516
Qy 741 PheTYRGLYAsnLeuProGlyMetAsnArgProValMetAlaSerLysAsnAspLysLeu 760
Db 2517 TTTGGGGTAACTTACCCGGAATGAC----- 2543
Qy 761 GlnLeuGlnAspCysLeuGlnPheSerArgThrAlaLysLeuLysValGlnThrIle 780
Db 2543 ----- 2543
Qy 781 ThrThrLysSerAsnSer1IleArgGlnGlyLysAsnGlnLeuPheProValValMetAsn 800
Db 2543 ----- 2543
Qy 801 GlyLysAspAspValLeuTYRProCysThrGlnLeuGlnArg1IlePheGlyPheProAlaHis 820
Db 2544 -----AGGATCTTGGGCTTCCCTGCTGAC 2567
Qy 821 TYRThrAspValSerAsnMetGlyArgGlyAlaArgGlnLysLeuLeuGlyArgSerTYR 840
Db 2568 TACACGAGACGTGTCAACATGTGGCGCGCGCGCGGTGAAGAGCTGTGGGACAGTCTCTGG 2627
Qy 841 SerValProVal1IleArgHisLeuPheAlaProLeuLysAspTYR-PheAlaCysGlu 859
Db 2628 AGTGTAACGGGTCTATGACACACTGTTTGCCCTTGAAAGACTACTTGTCTGTGAA 2684

```

```

RESULT 9
LOCUS AF151972 4149 bp mRNA linear ROD 08-JUN-2000
DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 4
ACCESSION AF151972
VERSION AF151972.1 GI:8347125

```

```

KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus.
REFERENCE
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
1 (bases 1 to 4149)
AUTHORS Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
Shen,Y.
TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing
isoforms in mouse embryonic tissue
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4149)
AUTHORS Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
Shen,Y.
TITLE Direct Submision
JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular
Biology, Institution of Basic Medical Sciences, Chinese Academy of
Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,
P.R.China
FEATURES
source Location/Qualifiers
1..4149
/organism="Mus musculus"
/strain="KM"
/db_xref="taxon:10090"
/dev_stage="8-9 day old embryo"
1..4149
/gene="Dnmt3b"
412..2802
/gene="Dnmt3b"
/function="methylates cytosine in DNA"
/note="alternatively spliced"
/codon_start=1
/product="DNA cytosine-specific methyltransferase isoform
4"
/protein_id="AA074518.1"
/db_xref="GI:8347126"
/translation="MKGDSRLNLEEAGYEBCIIVNGNFSQDSQSDTKDAPBPVLE
AICTEPVCTPRTGRSSRLSKREVSLNLYTODMGDRDDEVDGSDILMPK
LTREKDRTRTSRSPRATRSNGLSRORASPRITRGGQRHVOEYVEPRATR
SRRRASSASSTRSSPASVDPMEEVTPRSVSTPSVLSQGDQSDMDTQVDAESI
GDSIEYQDDKEFIDGLVWGRIKGFSPWMAVWSKATSKQAMPQMRWQVEGDKF
SEISADKLVALGLFSOHFNLAIFNLKLSYKAMVHTLEKARAGATFSSPGSLD
OLPMLNMAHGKPKTEIGLKPKNKOPVYNKSKVRSDSNLEPRRREKSRRTTN
DAAASESPPPKRLKTNSTYGGKDEDESERERASAEVTNNKNLEDCISCGKNPVS
FHPLEGGGLCOSGRDRELEFYMYDDEGYOSYCTVCCGRELILCSNTSCRCPCVEC
LELVAGAGTAEDAKLQBPWSCYMCIPORCHGVLRKRDMMNRLOPFTPDLEFEP
PKLYPAIPAKARKPRITRSLFDGIALTGVLKELGIKEKYIASVECAESIAVTVKH
EGQIKYVNDVRKTRTKNIEENGPFVDLVIGSPNDLSVNNPARKGLYEGGRLEPFEY
HLUNYTRPREGNRRPFPMFENVMKVDKDISRLACNPVMIDAIKVSAAHARARY
FWGNLPQNMRIREFGPAYHTDVSNMGRGAROKILGRSWSVPVIRHLPAPLKDYFACE"
BASE COUNT 1013 a 1082 c 1123 g 931 t
ORIGIN
Alignment Scores:
Pred. No.: 8.09e-228 length: 4149
Score: 4206.50 Matches: 794
Percent Similarity: 92.43% Conservative: 0
Best Local Similarity: 92.43% Mismatches: 2
Query Match: 91.64% Indels: 63
DB: 10 Gaps: 1
US-09-720-086-6 (1-859) x AF151972 (1-4149)
Qy 1 MetLysGlyAspSerArgHisLeuAsnGluGluGluValAlaSerGlyTYRGLuGluCys 20
Db 412 ATGAAGGAGACAGCAGACACTCTGAATGAAGAAAGAGGTGCCACGGGTATGAGAGTGC 471
Qy 21 IleLysValAsnGlyLysPheSerAspGlnSerAspThrLysAspAlaProSerPro 40
Db 472 ATTAATCGTTAATGGGAATTCAGTGAACGATCTCTGACACGAGAGGATGCTCCCTCACCC 531
Qy 41 ProValLeuGlnAla1IleCysThrGluProValCysThrProGluThrArgGlyArgArg 60

```



Db 532 CCAGCTTTGGAGGCAATCTGCACAGAGCCAGTCTGCACACAGAGAGCCAGAGCCGCGAGG 591  
Qy 61 SerSerArgLeuSerArgGluValSerSerLeuLeuAsnTyrThrGlnAspMet 80  
Db 592 TCAAGCTCCCGGCTCTTAAGAGGAGGTCTCCAGCTTCTGAAATACACGCGAGCATG 651  
Qy 81 ThrGlyAspGlyAspArgAspGluValAspAspGlyAsnGlySerAspIleLeuMet 100  
Db 652 ACAGAGATGAGACAGAGATGATGAAGTAGATGATGGAAATGGCTCTGATATTCTTAATG 711  
Qy 101 ProValLeuThrArgGluThrLysAspThrArgThrArgSerGluSerProAlaValArg 120  
Db 712 CCAAGCTCACCCGTGAGACCAAGACACAGGACGGCTCTGAAGCCCGGCTGTCCGA 771  
Qy 121 ThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnArgAlaSerProArgIleThr 140  
Db 772 ACCCGACATAGCAATGGGACCTCCAGCTTTGGAGAGCAAGAGCCCTCCCGAGAAATCAC 831  
Qy 141 ArgGlyArgGlnGlyArgHisValGlnGluTyrProValGluPheProAlaThrArg 160  
Db 832 CGAGGTCCGACAGGCGCCACCATGTGCAGGAGTACCTGTGGAGTTTCCGGCTTACAGG 891  
Qy 161 SerArgArgArgAlaSerSerSerAlaSerThrProThrProThrSerProAlaSerVal 180  
Db 892 TCTCGAGACGTCGAGCATCGTCTTCAGCAAGCACGCCATGGTGCATCCCTGCCAGCGTC 951  
Qy 181 AspPheMetGluGluValThrProLysSerValSerThrProSerValAspLeuSerGln 200  
Db 952 GACTTCATGGGAAGTAGTACACTTAAGAGCGTCACTACCCCATCAGTTGAGTTCGAGCAG 1011  
Qy 201 AspGlyAspGlnGluGlyMetAspThrGlnValAspAlaGluSerArgAspGlyAsp 220  
Db 1012 GATGAGATCAGAGAGGTATGATATCACACAGGTGATGCAGAGCATATATGGAGAC 1071  
Qy 221 SerThrGluTyrGlnAspAspLysGluPheGlyIleGlyAspLeuValTrpGlyLysIle 240  
Db 1072 AGCACAGATATCAGGATGATAAAGAGTTTGGAAATAGTGACCTCGTGTGGGGAAGATC 1131  
Qy 241 LysGlyPheSerTrpTrpProAlaMetValValSerTrpLysAlaThrSerLysArgGln 260  
Db 1132 AAGGCTTCTCTGTGGCTGGTCCATGGTGTGTCTCTGGAAAGCCACCTCCCAAGCAG 1191  
Qy 261 AlaMetProGlyMetArgTrpValGlnTrpPheGlyAspGlyLysPheSerGluIleSer 280  
Db 1192 GCCATGCCCGAATGGCTGGGTACAGTGGTTGGTATGGCAAGTTTCTGAGATCTCT 1251  
Qy 281 AlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsn 300  
Db 1252 GCTGACAACTGGTGGCTCTGGGCTGTTCCAGCCAGCACCTTTAATCTGGCTACCTTCAAT 1311  
Qy 301 LysLeuValSerTyrArgLysAlaMetTyrHisThrLeuGluLysAlaArgValArgAla 320  
Db 1312 AAGCTGGTTCTTATAGGAAGGCCATGTACACACTCTGGAGAAGCCAGGGTTCCAGCT 1371  
Qy 321 GlyLysThrPheSerSerProGlyGluSerLeuGluAspGlnLeuLysProMetLeu 340  
Db 1372 GGCAAGACCTTCTCCAGCAGTCTGGAGAGTCACTGGAGGACCCAGCTGAAGCCCATGCTG 1431  
Qy 341 GluTrpAlaHisGlyGlyPheLysProThrGlyLeuGlyLeuLysProAsnLysLys 360  
Db 1432 GAGTGGGCCACGGTGGCTTCAAGCCCTACTGGATTCAGGGCCCTCAACCCCAACAGAG 1491  
Qy 361 GlnProValValAsnLysSerLysValArgArgSerAspSerArgAsnLeuGluProArg 380  
Db 1492 CAACCAAGTGGTTAATAAGTCCAAAGGTGCTGCTTTCAGACAGTAGGAATTTAGAACCAGG 1551  
Qy 381 ArgArgGluAsnLysSerArgArgThrThrAsnAspSerAlaAlaSerGluSerPro 400  
Db 1552 AGACCGGAGACCAAAAGTCGAGAGCGCACCAACCAATGACTCTGTGCTTCTGAGTCCCC 1611  
Qy 401 ProProLysArgLeuLysThrAsnSerTyrGlyLysAspArgGlyGluAspGluGlu 420  
Db 1612 CCACCCAAAGCGCTCAAGACAAATAGCTATGGCGGGAAGGAGCCGAGGGGAGGATGAGGAG 1671

Qy 421 SerArgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAspArgCys 440  
Db 1672 AGCCGAGAACCGATGGCTTCTGAAGTACCACCAACAGGGCAATCTGGAGACCGCTGT 1731  
Qy 441 LeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyGlyLeuCys 460  
Db 1732 TTGTCCTGTGGAAAGNAGNACCTGTGTCTTCCACCCCTCTTTGAGSGTGGGCTGT 1791  
Qy 461 GlnSerCysArgAspArgPheLeuGluLeuPheTyrMetTyrAspGluAspGlyTyrGln 480  
Db 1792 CAGAGTTGCCGGGATCGCTTCTAGAGCTCTTCTATCATGTATGATGAGGACGGCTATCAG 1851  
Qy 481 SerTyrCysThrValCysCysGluGlyArgGluLeuLeuLeuCysSerAsnThrSerCys 500  
Db 1852 TCCTACTGCACCGTGTCTGTGAGGCGCGTGAACCTGCTGTGTGACGTAAACACCAAGTGC 1911  
Qy 501 CysArgCysPheCysValGluCysLeuGluValLeuValGlyAlaGlyThrAlaGluAsp 520  
Db 1912 TGCAGATGCTTCTGTGTGAGGTGTCTGGAGGTGTCTGTGGCGGCGAGCACAGCTGAGAT 1971  
Qy 521 AlaLysLeuGlnGluProTrpSerCysTyrMetCysLeuProGlnArgCysHisGlyVal 540  
Db 1972 GCCAAGCTGCAGGAAACCTTGGAGCTGTATATGTGCTCCTCAGCGCTGCATGGGTC 2031  
Qy 541 LeuArgArgArgLysAspTrpAsnMetArgLeuGlnAspPhePheThrThrAspProAsp 560  
Db 2032 CTTCCGACGCGAGAAAGATTGGAAACATGGCTCGCAAGCTTCTTCACTACTGATCTGTAC 2091  
Qy 561 LeuGluGluPheGluProProLysLeuTyrProAlaIleProAlaAlaLysArgArgPro 580  
Db 2092 CTGGAAGAAATTTGAGCCACCCAAAGTTGTACCAGCAATTCCTGCAGCCAAAGGAGGCC 2151  
Qy 581 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeu 600  
Db 2152 ATTAGAGTCTCTCTGTGTGATGGAAATTCGAAAGGAGTCTTGGTGTCTCAAGAGTTG 2211  
Qy 601 GlyIleLysValGluLysTyrIleAlaSerGluValCysAlaGluSerIleAlaValGly 620  
Db 2212 GGTATTAAAGTGGAAAGTACATTGGCTCCGAAAGTCTGTGCAGAGTCCATCGCTGTGGGA 2271  
Qy 621 ThrValLysHisGluGlyGlnIleLysTyrValAsnAspValArgLysIleThrLysLys 640  
Db 2272 ACTGTTAAGCATGAAGCCAGCATCAATATGTCAATCAGCTCCGGAATAATCACCAGAAA 2331  
Qy 641 AsnIleGluGluTrpGlyProPheAspLeuValIleGlyLysSerProCysAsnAspLeu 660  
Db 2332 AATATTGAAGAGTGGGCGCCGCTTCGACTTGGTGAATGGTGGAGCCCATGCAATGATCTC 2391  
Qy 661 SerAsnValAsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPhePheGlu 680  
Db 2392 TCTAACGTCAATCTGCCCGCAAGGTTTATATGAGGCGACAGGAAGGCTCTTCTTCGAG 2451  
Qy 681 PheTyrHisLeuLeuAsnTyrThrArgProLysGluGlyAspAsnArgProPhePheTyr 700  
Db 2452 TTTTACCACTTGTGAAATATACCCGCCCAAGGAGGCGCAACCCGTCATCTCTCTGG 2511  
Qy 701 MetPheGluAsnValValAlaMetLysValAsnAspLysLysAspIleSerArgPheLeu 720  
Db 2512 ATGTTCCAGAAATCTTGTGGCCATGAAGTGAATGACAGAAAGACATCTCAAGATTCCTG 2571  
Qy 721 AlaCysAsnProValMetIleAspAlaIleLysValSerAlaAlaHisArgAlaArgTyr 740  
Db 2572 GCATGTAAACCATGATGATCGATGCCATCAAGGTGTCTGTGCTCACAGGCGCCGGTAC 2631  
Qy 741 PheTrpGlyAsnLeuProGlyMetAsnArgProValMetAlaSerLysAsnAspLysLeu 760  
Db 2632 TTCTGGGGTAACCTTACCCGGAATGAAC----- 2658  
Qy 761 GluLeuGlnAspCysLeuGluPheSerArgThrAlaLysLeuLysLysValGlnThrIle 780  
Db 2658 ----- 2658

QY 781 ThrThrLysSerAsnSerIleArgGlnGlyLysAsnGlnLeuPheProValValMetAsn 800  
Db 2658 ----- 2658  
QY 801 GlyLysAspAspValLeuTrpCysThrGlnLeuGlnArgIlePheGlyPheProAlaHis 820  
Db 2659 ----- 2659  
QY 821 TyrThrAspValSerAsnMetGlyArgGlyAlaArgGlnLysLeuLeuGlnLysArgSerTrp 840  
Db 2683 TACACGGACGTTCTCAACATGGCCCGCCGCGCTGACAAAGCTGCTGAGCAGGCTCTGG 2742  
QY 841 SerValProValIleArgHisLeuPheAlaProLeuLysAspTyrPheAlaCysGln 859  
Db 2743 AGTGTACCGGTCATGACACACTGTTTCCCTTGAAGGACTTCTTCCCTGTGA 2799  
QY 10 AF068628 3946 bp mRNA linear ROD 06-DEC-1999  
LOCUS 28 AF068628  
DEFINITION Mus musculus DNA cytosine-5 methyltransferase 3B3 (Dnmt3b) mRNA,  
accession AF068628  
VERSION AF068628  
KEYWORDS AF068628.2 GI:6449473  
SOURCE Mus musculus.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
AUTHORS 1 (bases 1 to 3946)  
TITLE Okano, M., Xie, S., and Li, E.  
JOURNAL Cloning and characterization of a family of novel mammalian DNA  
MEDLINE (cytosine-5) methyltransferases  
PUBMED Nat. Genet. 19 (3), 219-220 (1998)  
9662389  
REFERENCE  
AUTHORS 2 (bases 1 to 3946)  
TITLE Xie, S., Okano, M., and Li, E.  
JOURNAL Direct Submission  
SUBMITTED (28-MAY-1998) CYRC, Mass. Gen. Hospital, 149 13th Street,  
Charlestown, MA 02129, USA  
3 (bases 1 to 3946)  
REFERENCE  
AUTHORS Okano, M., Chijiwa, T., Sasaki, H., and Li, E.  
TITLE Direct Submission  
SUBMITTED (04-NOV-1999) CYRC, Mass. Gen. Hospital, 149 13th Street,  
Charlestown, MA 02129, USA  
JOURNAL Sequence update by submitter  
COMMIT On Nov 18, 1999 this sequence version replaced gi:3327983.  
FEATURES  
Source  
1..3946  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/map="84.0 cM"  
1..3946  
/gene="Dnmt3b"  
1..3946  
/gene="Dnmt3b"  
/note="similar to EST sequences deposited in GenBank  
Accession Numbers AA116694, AA119979, AA172277, AA210568,  
AA407106, and AA575617".  
269..2599  
/gene="Dnmt3b"  
/function="de novo DNA methylation"  
/note="alternatively spliced product; contains Cys-rich  
region"  
/codon\_start=1  
/product="DNA cytosine-5 methyltransferase 3B3"  
/protein\_id="AAC40180.2"  
/db\_xref="GI:6449474"  
/translation="MKGDSRLNIEEGAGYECCITVNGNFSQSSDTKDAAPSPVLV  
AICTEPCTETRGRRSSRLSKREVSILNTYQDMTGDDGDDDEVDGSDILMPK  
LIRETKDTRRSRSPAVRTRHNGSTSLERORASPRITRGGRHHVQEVPEVPATR  
SRRRASSSASTPMSSPASVDMEETVPKSVSTPSVDLSQDGDGEMDTTQVDAESRD

GDSTETVODDKERFIDGLVMGKI KGRSWMVAMVUSWKAITSKROAMPGRWVFGDKP  
SEISADKVALGILFSQHFNATPNKLVSRKAMTYHTLEKAVRACKTFSSSPGESLED  
QLKPMLEMAHGGKPTGTEIGLRPNKQPNKSRRTTNDASAESPPPKRLKTNVYG  
KDRGDEBESRERMASEYETNNKGNLEDRCLSCGKKNPVSFHLFEGGLCOSCRDRLEL  
FYWVDEDEYQSYCTVCEGRELILCSNCSRCRCVCELELVGAGTAEPAKLOEPS  
CYMCLPQRCYGLRRKXDMNRLODPTTDPDLBEPPEPKLYPALPAARPRIVLSL  
FDGATGVLVLELGLKVEKYLASVCASIESIAGTVHGGQIKYVNDVKITKNIEE  
MGPELVIGSGSPCNDSLNNVNPARKGLYEGTGLFEEFYHLINLYTRPKEDNRPFFMMF  
ENVVAMKVNDDKDISRPLACNPMDALKVSAAHRAVFWGNLPGMNRIFGPAHYTD  
VSNMGRGAKRLGRSWSVPIRHLFALPKDVFACE"  
BASE COUNT 983 a 1006 c 1056 g 901 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3.7e-222 Length: 3946  
Score: 4105.50 Matches: 776  
Percent Similarity: 90.34% Conservative: 0  
Best Local Similarity: 90.34% Mismatches: 0  
Query Match: 89.44% Indels: 83  
DB: 10 Gaps: 2  
US-09-720-086-6 (1-859) x AF068628 (1-3946)  
QY 1 MetLysGlyAspSerArgHisLeuAsnGlnGlnGlyAlaSerGlyTyrGlnGlnCys 20  
Db 269 ATGAAGGAGACAGACGACACTCTTAATGAAGAAGGGTCCACGGGTATGAGAGTGC 328  
QY 21 IleIleValAsnGlyAsnPheSerAspGlnSerSerAspThrLysAspAlaProSerPro 40  
Db 329 ATTATCGTTAATGGGAACCTTCAGTGCAGTCTTCAGACAGAAAGATGCTCCTCACCC 388  
QY 41 ProValLeuGlnAlaIleCysThrGlnProValCysThrProGlnThrArgGlyArg 60  
Db 389 CCAAGTCTTGAGGCAATCGACAGAGCAGCTGCACACAGAGACAGAGCCGACAGG 448  
QY 61 SerSerSerArgLeuSerLysArgGlnValSerSerLeuLeuAsnTyrThrGlnAspMet 80  
Db 449 TCAAGCTCCCGGCTGTCTAAGAGGAGGTCTCCAGCTTCTGAATTACAGCAGCATG 508  
QY 81 ThrGlyAspGlyAspArgAspAspGlnValAspAspGlyAsnGlySerAspIleLeuMet 100  
Db 509 ACAAGAGATGAGACAGAGATGATGAAGTATGATGAGATGCTCTGATATTCTAATG 568  
QY 101 ProLysLeuThrArgGlnThrLysAspThrArgThrArgSerGlnSerProAlaValArg 120  
Db 569 CCAAAGCTCACCCGTCAGACCAAGAGACACGAGCGCTCTGAAGCCGCGCTGCCGA 628  
QY 121 ThrArgHisSerAsnGlyThrSerSerLeuGlnArgGlnArgAlaSerProArgIleThr 140  
Db 629 ACCGACATAGCAATGAGACTCCAGCTTGAGAGGCAAGAGCTCCGCCAGATCAC 688  
QY 141 ArgGlyArgGlnGlyArgHisIleValGlnGlnTyrProValGluPheProAlaThrArg 160  
Db 689 CGAGGTCCGACAGGCCGCCACCATGTGACAGAGTACCTGTGAGATTCCGCTACAGG 748  
QY 161 SerArgArgArgArgAlaSerSerSerAlaSerThrProTyrSerSerProAlaSerVal 180  
Db 749 TCTCGAGACGTCGAGCATGCTCTTACAGCAAGCACCCATGATCATCCCTCCAGGTC 808  
QY 181 AspPheMetGlnGlnValThrProLysSerValSerThrProSerValAspLeuSerGln 200  
Db 809 GACTTCATGGAAGAAGTACACCTTAAGACCGTCAGTACCATCAGTTCGATGAGCCAG 868  
QY 201 AspGlyAspGlnGlnGlnMetAspThrThrGlnValAspAlaGlnSerArgAspGlyAsp 220  
Db 869 GATGAGATCAGAGAGGTATGATACCAACAGGTGATCCAGAGCAGAGAGATGAGGAC 928  
QY 221 SerThrGlnTyrGlnAspAspLysGluPheGlyIleGlyAspLeuValTrrGlyLysIle 240  
Db 929 AGCAGAGATACAGATGATTAAGACTTTGGAATAGTGACCTCGTGTGGGAAAGATC 988  
QY 241 LysGlyPheSerTyrTrpProAlaMetValValSerTrpLysAlaThrSerLysArgGln 260

Db 989 AAGGGCTTCTCTGGTGGCTGGCATGGTGTCTCTGGAAGCCACCTCCACGACAG 1048  
Qy 261 AlaMetProGlyMetArgTrpValGlnTrpPheGlyAspGlyLysPheSerGluIleSer 280  
Db 1049 GCCATGCCCGGAATGGCTGGGTACAGTGGTTGGTGATGGCAAGTTTCTGAGATCTCT 1108  
Qy 281 AlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsn 300  
Db 1109 GCTGACAACTCGTGGCTCTGGGGCTGTTGAGCAGCACCTTTAATCTGGCTACCTTCAAT 1168  
Qy 301 LysLeuValSerTyrArgLysAlaMetTyrHisThrLeuGluLysAlaAtrqValArgAla 320  
Db 1169 AAGCTGGTTCTTATAGGAAGGCCATGTACACACCTCTGGAGAAAGCCAGGGTTCGAGCT 1228  
Qy 321 GlyLysThrPheSerSerSerProGlyGluSerLeuGluAspGlnLeuLysProMetLeu 340  
Db 1229 GGCAAGACCTTCTCAGACAGTCTGGAGAGTCACTGGAGACCAAGCTGAAGCCCATGCTG 1288  
Qy 341 GluTrpAlaHisGlyGlyPheLysProThrGlyLeuGlyLeuLysProAsnLysLys 360  
Db 1289 GAGTGGGCCACCGTGGCTTCAAGCCCTACTGGATCGAGGGCTCAAAACCCCAAGAAG 1348  
Qy 361 GlnProValValAsnLysSerLysValArgArgSerAspSerArgAsnLeuGluProArg 380  
Db 1349 CAACCA----- 1354  
Qy 381 ArgArgGluAsnLysSerArgArgArgThrThrAsnAspSerAlaAlaSerGluSerPro 400  
Db 1355 -----GAGAACAAAGTCCAGACGACACCAACCAATGACTCTGCTGCTTCTGAGTCCGCC 1408  
Qy 401 ProProLysArgLeuLysThrAsnSerTyrGlyGlyLysAspArgGlyGluAspGluGlu 420  
Db 1409 CCACCCAGCGCTCAAGACAAATAGCTATGGCGGGAAGGACCGAGGGGAGATGAGGAG 1468  
Qy 421 SerArgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAspArgCys 440  
Db 1469 AGCCGAGAACGGATGGCTTCTGAAGTCAACCAACAAAGGGCAATCTGGAAGACCGCTGT 1528  
Qy 441 LeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyGlyLeuCys 460  
Db 1529 TTGTCTGTGGAAAGAAAGAACCTGTGTCTTCCACCCCTCTTTGAGGGTGGGCTCTGT 1588  
Qy 461 GlnSerCysArgAspArgPheLeuGluLeuPheTyrMetTyrAspGluAspGlyTyrGln 480  
Db 1589 CAGAGTTCCGGGATCGCTTCTTAGAGCTCTTCTACATGTATGTATGAGGACGGCTATCAG 1648  
Qy 481 SerTyrCysThrValCysCysGluGlyArgGluLeuLeuLeuLeuCysSerAsnThrSerCys 500  
Db 1649 TCCTACTGCACCGTGTGTGTGAGGGCCGTGAACCTGTCTGTCTGTCTGACGTAAACACAGCTGC 1708  
Qy 501 CysArgCysPheCysValGluCysLeuGluValLeuValGlyAlaGlyThrAlaGluAsp 520  
Db 1709 TGCAGATGCTTCTGTGTGGAGTGTCTGGAGTGTCTGGGCGGAGGACACAGCTGAGGAT 1768  
Qy 521 AlaLysLeuGlnGluProTrpSerCysTyrMetCysLeuProGlnArgCysHisGlyVal 540  
Db 1769 GCCAAGCTGCAGGAACCCCTGGAGCTGTATATGTGCTCTCCCTCAGCGCTGCCATGGGGTC 1828  
Qy 541 LeuArgArgArgLysAspTrpAsnMetArgLeuGlnAspPhePheThrThrAspProAsp 560  
Db 1829 CTCACGCGCAGGAAGATTGGAACATGCGGCTGTCAGACCTTCTTCACTACTGATCTGAC 1888  
Qy 561 LeuGluGluPheGluProProLysLeuTyrProAlaIleProAlaAlaLysArgArgPro 580  
Db 1889 CTGGAAGAAATTTAGGCCACCAAGTTGTACCCCAATTTCTGACGCAAAAGAGGCC 1948  
Qy 581 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeu 600  
Db 1949 ATTAGAGTCTCTCTCTGTTTGTATGGAATTTGCAACGGGGTACTTTGGTCTCAAGAGTTG 2008  
Qy 601 GlyIleLysValGluLysTyrIleAlaSerGluValCysAlaGluSerIleAlaValGly 620  
Db 2009 GGATTAAGTGGAAAGTACATTGCTCCGGAAGTCTGTGCAGAGTCCATCGTCTGGGA 2068

Qy 621 ThrValLysHisGluGlyGlnIleLysTyrValAsnAspValArgLysIleThrLysLys 640  
Db 2069 ACTGTTAAGCATGAAGCGCCAGATCAATATGTCAATGACGTCGCCGAAAATCACCAGAAA 2128  
Qy 641 AsnIleGluGluTrpGlyProPheAspLeuValIleGlyGlySerProCysAsnAspLeu 660  
Db 2129 AATATTGAAGAGTGGGGCCGTTTCGACTTGGTGAATGGTGAAGCCCATGCAATGATCTC 2188  
Qy 661 SerAsnValAsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPhePheGlu 680  
Db 2189 TCTAACGTCATCTGCGCCGCAAGGTTTATATGAGGGCACAGGAAGGCTCTTCTTCGAG 2248  
Qy 681 PheTyrHisLeuLeuAsnTyrThrArgProLysGluGlyAspAsnAspProPheThrTrp 700  
Db 2249 TTTTACCATTTGCTGAATATATCCGCCCAAGAGGGCGCAACCCGTCATCTCTCTGG 2308  
Qy 701 MetPheGluAsnValValAlaMetLysValAsnAspLysLysAspIleSerArgPheLeu 720  
Db 2309 ATGTTCCAGAAATGTTGTGGCCATGAAGTGAATGACAGGAAGACATCTCAAGATTCTCTG 2368  
Qy 721 AlaCysAsnProValMetIleAspAlaIleLysValSerAlaAlaHisArgAlaArgTyr 740  
Db 2369 GCATGTAAACCCAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2428  
Qy 741 PheTyrGlyAsnLeuProGlyMetAsnArgProValMetAlaSerLysAsnAspLysLeu 760  
Db 2429 TTCTGGGGTAACCTACCCCGAATGAAC----- 2455  
Qy 761 GluLeuGlnAspCysLeuGluPheSerArgThrAlaLysLeuLysLysValGlnThrIle 780  
Db 2455 ----- 2455  
Qy 781 ThrThrLysSerAsnSerIleArgGlnGlyLysAsnGlnLeuPheProValValMetAsn 800  
Db 2455 ----- 2455  
Qy 801 GlyLysAspAspValLeuTrpCysThrGluLeuGluArgIlePheGlyPheProAlaHis 820  
Db 2456 -----AGGATCTTCCGGCTTCCCTGCTCAC 2479  
Qy 821 TyrThrAspValSerAsnMetGlyArgGlyValAlaArgGlnLysLeuLysLeuGlyArgSerTrp 840  
Db 2480 TACACGAGCTGTCCACATGCG 2539  
Qy 841 SerValProValIleArgHisLeuPheAlaProLeuLysAspTyrPheAlaCysGlu 859  
Db 2540 AGTGTACCGGTCTACAGACACCTGTTTGGCCCTTGAAGGACTACTTTGCTGCTGAA 2596  
RESULT 11  
AF151976 3974 bp mRNA linear ROD 08-JUN-2000  
LOCUS Mus musculus DNA cytosine-specific methyltransferase isoform 8  
DEFINITION (Dnmt3b) mRNA, complete cds.  
ACCESSION AF151976  
VERSION AF151976.1 GI:8347136  
KEYWORDS Mus musculus.  
SOURCE Mus musculus.  
ORGANISM Mus musculus.  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS 1 (bases 1 to 3974)  
Yin, B., Chen, Y.T., Zhu, M., Luo, Y.J., Zhu, N., Xu, S.C., Wu, G.Y. and  
Shen, Y.  
TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing  
isoforms in mouse embryonic tissue  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3974)  
Yin, B., Chen, Y.T., Zhu, M., Luo, Y.J., Zhu, N., Xu, S.C., Wu, G.Y. and  
Shen, Y.  
TITLE Direct Submission  
JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular  
Biology, Institution of Basic Medical Sciences, Chinese Academy of

Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,  
P.R.China

FEATURES  
source

Location/Qualifiers  
1. .3974  
/organism="Mus musculus"  
/strain="KM"  
/db\_xref="taxon:10090"  
/dev\_stage="8-9 day old embryo"  
1. .3974  
/gene="Dmmt3b"  
297. .2627  
/gene="Dmmt3b"  
/function="methylates cytosine in DNA"  
/note="alternatively spliced"  
/codon\_start=1  
/product="DNA cytosine-specific methyltransferase isoform 8"

gene

CDS

/protein\_id="AA074522.1"  
/db\_xref="GI:8347137"  
/translation="MKGDSRLHNEEBGAGYEECTIVNGNFSDQSSDTKADPSPVLE  
AICTEPVCPTETRRSSRLSKREYSILNTYQDMTGDDDEVDGSGSDILMPK  
LRTETDPTRRSEPAVTRHNGTSLERORABPRIHGRGRHHVOEYVPEEPATR  
SRRRASSASTPWSPASVDPEBETPKSVSTPSVDLSQDDQEGMDTTPVDASIT  
GDSITQDDKEFGIDLVWGKIKGFSWPMAVWSKATSKRAMPMRWQVFGDKF  
SEISADKLVALGIFSOHNLATFNKLVSYRKAMVHTLKKARYAGATPSSPGESLED  
OLKPMLEMAHAGFPETGLEKPNKOPENKSRRTTDSASESPPKLKTNSYGG  
KDRGDESERBMASEVTNNKNLEDRCLSCGKNPVPFPIPEGGLCGCRDFLEL  
FMYDEBDGYOSTCVCCEGRELICNTSCRCFCEVECELVAGTAEDAKLOEWS  
CYMCLPQRCHGVARRKMMMLQDFETDDEFEPEPKLYPAIPAAKRPPIKVLSTL  
FDGIATGYLVKELGKYEKYLASEVCAESIAGVTKEGQIKYVDRITKKNIES  
MGPFDIVGSSPNDLSNVPARKGLYEGTRGALPFEFPHLNTTRKEGDNRPFFMWF  
ENVVAKVNDKXDISRLACNPVMDAIKVSAAHARARYFMGNLPGMNRIFFPHAYTD  
VSNMGRGAKQLGRSWSVPVIRHLPALPKDYFACE"

BASE COUNT 972 a 1032 c 1078 g 892 t

ORIGIN

Alignment Scores:

Pred. No.: 3.38e-221 Length: 3974  
Score: 4088.50 Matches: 774  
Percent Similarity: 90.10% Conservative: 0  
Best Local Similarity: 90.10% Mismatches: 2  
Query Match: 89.07% Indels: 83  
DB: 10 Gaps: 2

US 720-086-6 (1-859) x AF151976 (1-3974)

Oy 1 MetLysGlyAspSerArgHisLeuAsnGluGluGluValAsnSerGlyTyrGluGluCys 20  
Db 297 ATGAAGGAGACAGACGACATCGAATGAAGAAGAGGCTGCCAGCGGTATGAGAGAGTGC 356  
Oy 21 IleIleValAsnGlyAsnPheserAspGlnserSerAspThrIleAspAlaProserPro 40  
Db 357 ATTATCGTTAATGGAACTTCAGTACCAAGTCTCAGACAGAGAGATCTCCCTCACCC 416  
Oy 41 ProValIleuGluAlaIleCysThrGluProValCysThrProGluThrArgGlyArgArg 60  
Db 417 CCAGTCTTGAGGCAATCTGCACAGCAAGTGTCAACCAAGAGACCAAGAGCGCGAGG 476  
Oy 61 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuAsnTyrThrGlnAsnMet 80  
Db 477 TCAAGCTCCCGCTGCTTAAAGAGGAGGTCTCCAGCTTCTGAATTACACGACAGCATG 536  
Oy 81 ThrGlyAspGlyAspArgAspGluValAspAspGlyAsnGlySerAspIleLeuMet 100  
Db 537 ACAGGAGATGGAGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 596  
Oy 101 ProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSerProAlaValArg 120  
Db 597 CCAAGCTCACCCTGAGACCAAGACACGAGCGCTCTGAAGCCCGGCTGTCCGA 656  
Oy 121 ThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnArgAlaSerProArgIleThr 140  
Db 657 ACCGACATAGCAATGGAGACTCCAGCTTGAGAGGAGCAAGAGCTCCCAAGATATACC 716

Oy 141 ArgGlyArgGlnGlyArgHisIleValGlnGluTyrProValGluPheProAlaThrArg 160  
Db 717 CGAGTCGGCAGGCGCGCCACCATGTGCAGAGTACCTTGAGATTCCGGCTACAGG 776  
Oy 161 SerArgArgArgArgAlaSerSerSerLysSerThrProTyrSerSerProAlaSerVal 180  
Db 777 TCTCGGAGACGTGAGCATCTGCTTCAAGCAAGCACCGCATGATATCCCTCCAGCGTC 836  
Oy 181 AspPheMetGluGluValThrProLysSerValSerThrProSerValAspLeuSerGln 200  
Db 837 GACTTCATGGAAGATGACACCTTAAGACCGTCAGTACCCCATCATGATGATGAGCCAG 896  
Oy 201 AspGlyAspGlnGluGlyMetAspThrThrGlnValAspAlaGluSerArgAspGlyAsp 220  
Db 897 GATGAGATCAGAGAGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 956  
Oy 221 SerThrGluTyrGluAspAspLysGluPheGlyIleGlyAspLeuValTyrGlyLysIle 240  
Db 957 AGCAGAGATACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1016  
Oy 241 LysGlyPheSerThrProAlaMetValValSerTyrLysAlaThrSerLysArgGln 260  
Db 1017 AAGGCTTCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1076  
Oy 261 AlMetProGluMetArgTyrValGluTyrPheGlyAspGlyLysPheSerGluIleSer 280  
Db 1077 GCCATGCCCCGAATGGCTGGGATGATGATGATGATGATGATGATGATGATGATGATG 1136  
Oy 281 AlaAspLysLeuValAlaLeuGlyLysPheSerGlnHisPheAsnLeuAlaThrPheAsn 300  
Db 1137 GCTGACCAACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1196  
Oy 301 LysLeuValSerTyrArgLysAlaMetTyrHisThrLeuGluLysAlaArgValArgAla 320  
Db 1197 AAGCTGTTCTTATAGGAAGGACATGATCAACATCTGAGAAAGCCAGGGTTCGAGCT 1256  
Oy 321 GlyLysThrPheSerSerSerProGlyValLysLeuGluAspGlnLeuLysProMetLeu 340  
Db 1257 GGCAAGACCTTCTCCAGCACTCTCGAGAGTCACTGAGAGACCAAGCTGAAGCCCATGCTG 1316  
Oy 341 GlnTyrPheIleAsnGlyValPheLysProThrGlyIleGluGlyLysLeuLysProAsnLys 360  
Db 1317 GAGTGGGCCACAGGAGGCTTCAACCTTACTGGGATGAGGGCTTCAACCCCAAGAAAG 1376  
Oy 361 GlnProValValAsnLysSerLysValArgArgSerAspSerArgAsnLeuGluProArg 380  
Db 1377 CAACCC----- 1382  
Oy 381 ArgArgGluAsnLysSerArgArgArgThrThrAsnAspSerAlaIleSerGluSerPro 400  
Db 1383 -----GAGAACAAAGCTCGAAGACGACACCAATGACTCTGCTCTTGTGATCCCCC 1436  
Oy 401 ProProLysArgLeuLysThrAsnSerTyrGlyLysAspArgGlyGluAspGluGlu 420  
Db 1437 CCACCCAGCCGCTCAAGACAAATAGCTTATGCGGGAAGACCAAGGGAGATGAGAGG 1496  
Oy 421 SerArgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAspArgCys 440  
Db 1497 AGCCGAGAACGATGCTTGTGAAGTCAACCAACAGGGAATCTTGGAAGACCGCTGT 1556  
Oy 441 LeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyLysCys 460  
Db 1557 TTGTCTCTGGAAGAAAGAACCTGTGTCTTCAACCCCTCTTGAAGGTGGCTGTGT 1616  
Oy 461 GlnSerCysArgAspArgPheLeuGluLysPheTyrMetTyrAspGluAspGlyTyrGln 480  
Db 1617 CAGAGTCCCGGATGCTTCTTGAAGCTCTTCAATGATGATGATGATGATGATGATGATG 1676  
Oy 481 SerTyrCysThrValCysCysGluGlyArgGluLeuLeuLeuCysSerAsnThrSerCys 500  
Db 1677 TCCTACTGACCGGTGTGTGTGAGGGCGGTGAAGTCTGCTGTGCTGCTGCTGCTGCTGCTG 1736

```
QY 501 CysArgCysPheCysValGluCysLeuGluValLeuValGlyAlaGlyThrAlaGluAseP 520
Db 1737 TGCAGATGCTTCTGTGTGGAGTGTCTGGAGGTGCTGTGGCGCAGGCACAGCTGAGGAT 1796
QY 521 AlaLysLeuGluGluProTrpSerCysTyrMetCysLeuProGlnArgCysHisGlyVal 540
Db 1797 GCCAAGTGCAGAACCCCTGGAGCTGCTATATGTGCTCTCCCTCAGCGCTGCATGGGTC 1856
QY 541 LeuArgArgLysAspTrpAsnMetArgLeuGlnAspPhePheThrThrAspProAsp 560
Db 1857 CTCCGACGAGGAAAGATTGGAACTGCGCTGCAAGCTTCTTCACTACCTGATCTCTGAC 1916
QY 561 LeuGluGluPheGluProProLysLeuTyrProAlaIleProAlaAlaLysArgArgPro 580
Db 1917 CTGGAGAGATTGTGACCACTTGTACCCAGCAATCTCTGAGCCAAAGAGGCC 1976
QY 581 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeu 600
Db 1977 ATTAGAGTCTCTCTCTGTTGATGGAATTGCAACGGGTACTTGGTGTCTCAAGAGATTG 2036
QY 601 GlyIleValGluLysTyrIleAlaSerGluValCysAlaGluSerIleAlaValGly 620
Db 2037 GGTATTAAAGTGGAAAGTACATTGCTCTCGAAGTCTGTGCAGAGTCCATCGCTGTGGA 2096
QY 621 ThrValLysHisGluGluGlnIleLysTyrValAsnAspValArgLysIleThrLysLys 640
Db 2097 ACTGTTAAGCATGAAGGCCAGATCAATATGTCAATGACGTCCGGAATAATCACCAGAAA 2156
QY 641 AsnIleGluGluTrpGlyProPheAspLeuValIleGlyGlySerProCysAsnAspLeu 660
Db 2157 AATAATTGAAGAGTGGGGCCGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2216
QY 661 SerAsnValAsnProAlaArgLysGlyLeuTyrGluClyThrGlyArgLeuPhePheGlu 680
Db 2217 TCTAACGTCAATCTCTCCGCAAGGTTTATATAGAGGACAGAGAGCTCTTCTTCGAG 2276
QY 681 PheTyrHisLeuLeuAsnTyrThrArgProLysGluGlyAspAsnArgProPhePheTrp 700
Db 2277 TTTTACCATTGCTGAATATATACCCGCCCAAGAGGCGGACAAACCGTCTCTTCTTCTGG 2336
QY 701 MetPheGluAsnValValAlaMetLysValAsnAspLysLysAspIleSerArgPheLeu 720
Db 2337 ATGTTTCGAGAAATGTTGTGCCATGAAGTGAATGACAGAAAGACATCTCAAGATTCTGT 2396
QY 721 AlaCysAsnProValMetIleAspAlaIleLysValSerAlaAlaHisArgAlaArgTyr 740
Db 2397 GCATGTAACCCAGTGTATGATCGATCCATCAAGGTGTCTGCTGCTCAGGCGCCGCTAC 2456
QY 741 PheTrpGlyAsnLeuProGlyMetAsnArgProValMetAlaSerLysAsnAspLysLeu 760
Db 2457 TTCTGGGTAACTACCCGGAATGAAC----- 2483
QY 761 GluLeuGlnAspCysLeuGluPheSerArgThrAlaLysLeuLysLysValGlnThrIle 780
Db 2483 ----- 2483
QY 781 ThrThrLysSerAsnSerIleArgGlnGlyLysAsnGlnLeuPheProValValMetAsn 800
Db 2483 ----- 2483
QY 801 GlyLysAspValLeuTrpCysThrGluLeuGluArgIlePheGlyPheProAlaHis 820
Db 2484 -----AGGATCTTCGGCTTCCCTGCTCTC 2507
QY 821 TyrThrAspValSerAsnMetGlyArgGlyAlaArgGlnLysLeuLeuGlyArgSerTrp 840
Db 2508 TACACGGAGCTGTCCAACTGGCGCGCGCGCTGAGAGCTGTCTGCGGAGGCTCTCTGG 2567
QY 841 SerValProValIleArgHisLeuPheAlaProLeuLysAspTyrPheAlaCysGlu 859
Db 2568 AGTGTAACCGGTATCATCAGACACCTGTTGTCCTTGGAGGACTACTTGTGCTGTGAA 2624
```

RESULT 12

```
AF151971 4089 bp mRNA linear ROD 08-JUN-2000
LOCUS Mus musculus DNA cytosine-specific methyltransferase isoform 3
DEFINITION (Dnmt3b) mRNA, complete cds.
ACCESSION AF151971
VERSION AF151971.1 GI:8347122
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4089)
AUTHORS Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
Shen,Y.
TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing
isoforms in mouse embryonic tissue
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4089)
AUTHORS Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
Shen,Y.
DIRECT SUBMISSION
SUBMITTED (17-MAY-1999) Department of Biochemistry and Molecular
Biology, Institution of Basic Medical Sciences, Chinese Academy of
Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,
P.R.China
FEATURES
source
1..4089
/organism="Mus musculus"
/strain="KM"
/db_xref="taxon:10090"
/dev_stage="8-9 day old embryo"
1..4089
/gene="Dnmt3b"
412..2742
/gene="Dnmt3b"
/function="methylates cytosine in DNA"
/notes="alternatively spliced"
/codon_start=1
/product="DNA cytosine-specific methyltransferase isoform
3"
/protein_id="AAF74517.1"
/db_xref="GI:8347123"
/translation="MKGDSRHLNEERGASGYECIIVNGNFSDOSSDTKADSPVPVLE
ALCTEFCVETGRKSSKLSKREVSLLNYTQDMTGDGDRDDEVDNGSDILMP
LRTETKDTRESAPVTRTHNSGTSLERQASPRITRGQRHHVQEIYVPEFPA
SRRRASASASTPWSVSDVDFMEVTPKSVSPVSDLSODGDEGMDTTQVDAESL
GSTTYQDDKEFGIGLVNKGIFGFSWPMVSVKATSKRQAMPKRWQVGFQDKF
SBISADKLVALGFSQHLNFNLATNKLVSRYKAMYHTLEKARVAGKTFSSPSLEL
QKPMLEWAHGGFKPTGIEGLKPKNKQPKNSRRRTTNDASAPPPKRLKTNVGG
KDRGDEESERMASVETNNKGNLEDRCLSCGKNPVSEFHLFEGGLCQSCDRLEL
FYMYDEGYSYCTCCGRELCLSCNCTSCRCFCVCEVLVAGTAGTADAKLQSPWS
CYMCLPQRCGVLRERKDMNRLQDRFTTDPDLEEFEPKLYPAIPAKRRPIRLSL
FQGIATGILVLKELGKVEKIASVEVCAESIAGTVKHEGQIKVNDVRKTKKNIEE
WGFPLVIGGSPCNDLUSNVNPAKGLYEGTGRUFFEYFHLNLYTRPKSGDNRPFFWMP
ENVMAMVNDKDDISRLACNPVMDAIVSAAHARYFNGNLPGMNRIFGPPAHYTD
VSNMGRGAROKLIGRSWSPVIRHLFAPLKDYPACE"
BASE COUNT 995 a 1071 c 1104 g 919 t
ORIGIN
Alignment Scores:
Pred. No.: 3 5e-221 Length: 4089
Score: 4088.50 Matches: 774
Percent Similarity: 90.10% Conservative: 0
Best Local Similarity: 90.10% Mismatches: 2
Query Match: 89.07% Indels: 83
DB: 10 Gaps: 2
US-09-720-086-6 (1-859) x AF151971 (1-4089)
QY 1 MatLysGlyAspSerArgHisLeuAsnGluGluGlyAlaSerGlyTyrGluGluCys 20
Db 412 ATGAGGGAGACAGACACATCTGAATGAGAGAGGGTCCAGCGGTATGAGGAGTGC 471
```

QY 21 IleIleValAsnGlyAsnPheserAspGlnSerSerAspThrIysAspAlaProSerPro 40  
 DB 472 ATTATCGTTAAATGGGAATCTTCAATGACCACTCTCAGACAGAAAGATGCTCCCTCACC 531  
 QY 41 ProValIleuGlnAlaIleCysThrGluProValCysThrProGluThrArgGlyArg 60  
 DB 532 CCACTCTTGGAGCAATCTGCAACAGACCAAGTGTGCAACCAAGACCAAGAGCCGACAG 591  
 QY 61 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuAsnTyrThrGlnAspMet 80  
 DB 592 TCAAGCTCCGGCTGCTTAAGAGGAGGTCTCCAGCTTCTGAATTACCGCAGGACATG 651  
 QY 81 ThrGlyAspGlyAspArgAspAspGluValAspAspGlyAsnGlySerAspIleLeuMet 100  
 DB 652 ACAGAGAGATGGAGACAGATGATGATGATGATGATGATGATGATGATGATGATGATG 711  
 QY 101 ProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSerProAlaValArg 120  
 DB 712 CCAAGCTCACCCTGAGACCAAGACCAAGACCGCTCTGAAAGCCGCGCTGCCA 771  
 QY 121 ThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnArgAlaSerProArgIleThr 140  
 DB 772 ACCGACATAGCAATGGGACCTCCAGCTTGGAGAGGCAAAAGCCTCCCAAGATACAC 831  
 QY 141 ArgGlyArgGlnGlyArgHisIleValGlnGlyProValGluPheProAlaThrArg 160  
 DB 832 CGAGGTCCGACAGGCGCCGCCACATGTCAGAGATACCTGTGAGATTCCGGCTACCAAG 891  
 QY 161 SerArgArgArgArgAlaSerSerSerAlaSerThrProTyrSerSerProAlaSerVal 180  
 DB 892 TCTGGAGACATCGACATGCTCTTACAGCAAGCAGCAGCATGCTATCCCTGCCAGGCTC 951  
 QY 181 AspPheMetGluGluValThrProLysSerValSerThrProSerValAspLeuSerGln 200  
 DB 952 GACTTCATGGAAGAGAGACGACCTTAAGACGCTCAGTACCCCATCAGTGAATTGAGCCAG 1011  
 QY 201 AspGlyAspGlnGluGlyMetAspThrThrGlnValAspAlaGluSerArgAspGlyAsp 220  
 DB 1012 GATGAGATCAGAGAGGTATGATACACACAGGTGATGACAGACATATATGAGAC 1071  
 QY 221 SerThrGluTyrGlnAspAspLysGluPheGlyIleGlyAspLeuValTyrGlyLysIle 240  
 DB 1072 AGCACAGATACAGATGATGAATAAGATTGGAATAGTGACCTCTGTGGGAAAGATC 1131  
 QY 241 LysGlyPheSerTyrTyrProAlaMetValIleSerTyrLysAlaThrSerLysArgGln 260  
 DB 1132 AAGGCTTCTCTGCTGCTGCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1191  
 QY 261 AlaMetProGlyMetArgTyrValGlnTyrPheGlyAspGlyLysPheSerGluIleSer 280  
 DB 1192 GCCATGCCCGGAATGCGCTGGGTACAGTGGTTGGTATGGCAAGTTTCTTGATCTCT 1251  
 QY 281 AlaAspLysLeuValAlaLeuGlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsn 300  
 DB 1252 GCTGACAAACTGGTGGCTGCTGGGCTTCCAGCCAGCACTTTAATCTGGTACTCTCAT 1311  
 QY 301 LysLeuValSerTyrArgLysAlaMetTyrHisThrLeuGluLysAlaArgValArgAla 320  
 DB 1312 AAGCTGTTCTTATAGAAAGCCATGTACCACTCTGAGAAAGCCAGGGTTCCAGCT 1371  
 QY 321 GlyLysThrPheSerSerSerProGlyGluSerLeuGluAspGlnLeuLysProMetLeu 340  
 DB 1372 GCGAAGACTTCTCCACAGCTCTGAGAGTCACTGAGAGACCACTGAAGCCATGCTG 1431  
 QY 341 GluTyrAlaHisGlyGlyPheLysProThrGlyIleGluGlyLeuLysProAsnLysLys 360  
 DB 1432 GAGTGGGCGCCACGCTGCTTCAAGCTTACTGAGATCGAGGGCTCAAAACCAACAAAG 1491  
 QY 361 GlnProValValAsnLysSerLysValArgArgSerAspSerArgAsnLeuGluProArg 380  
 DB 1492 CAACCA----- 1497  
 QY 381 ArgArgGluAsnLysSerArgArgArgThrThrAsnAspSerAlaAlaSerGluSerPro 400

DB 1498 -----GAGCAACAAAGTCGAAGCGCACACCAATGACTCTGCTGCTTCTGAGTCCCC 1551  
 QY 401 ProProLysArgLeuLysThrAsnSerTyrGlyGlyLysAspArgGlyGluAspGluGlu 420  
 DB 1552 CCAACCAAGCGCTCAAGACAAATAGCTATGGCGGGAAGGACCAAGGAGATGAGAGAG 1611  
 QY 421 SerArgGluArgMetAlaSerGluValThrAsnAsnLysGlyAsnLeuGluAspArgCys 440  
 DB 1612 AGCCGAGAACGATGCTCTTGAAGTCAACCAACAGGCAATCTGGAAGACCGCTGT 1671  
 QY 441 LeuSerCysGlyLysLysAsnProValSerPheHisProLeuPheGluGlyLeuCys 460  
 DB 1672 TTGTCTGTGAAAGAAAGAACCTGTGTCTTCCACCCCTCTTGAAGGTGGCTGTGT 1731  
 QY 461 GlnSerCysArgAspArgPheLeuGluLeuPheTyrMetTyrAspGluAspGlyTyrGln 480  
 DB 1732 CAGAGTCCCGGGATGCTCTTGAAGTCAACCAACAGGCAATCTGGAAGACCGCTGT 1791  
 QY 481 SerTyrCysThrValCysCysGluGlyArgGluLeuLeuLeuCysSerAsnThrSerCys 500  
 DB 1792 TCCCTACTGCACCGGTGTGTGTGAGAGCGGTGAACCTGCTGTGTGAGTAACACAGCTGC 1851  
 QY 501 CysArgCysPheCysValGluCysLeuGluValLeuValGlyAlaGlyThrAlaGluAsp 520  
 DB 1852 TGCAGATCTCTGT 1911  
 QY 521 AlaLysLeuGlnGluProTyrSerCysTyrMetCysLeuProGlnArgCysHisGlyVal 540  
 DB 1912 GCCAAGCTGCAGAAACCTGTGAGTGTATATGTGCTCTCCACGCGCTGCCATGGGGTC 1971  
 QY 541 LeuArgArgArgLysAspTyrPheMetArgLeuGlnAspPheThrThrAspProAsp 560  
 DB 1972 CTCGACCGCAGAAACCTGTGAGTGTATATGTGCTCTCCACGCGCTGCCATGGGGTC 2031  
 QY 561 LeuGluGluPheGluProProLysLeuTyrProAlaIleProAlaAlaLysArgArgPro 580  
 DB 2032 CTGGAAGATTTGAGCCACCAAGTTGATCCAGCAATCTCTGACCCAAAGAGAGGCC 2091  
 QY 581 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeu 600  
 DB 2092 ATTAGATCTCTGCTGT 2151  
 QY 601 GlyIleLysValGluLysTyrIleAlaSerGluValCysAlaGluSerIleAlaValGly 620  
 DB 2152 GGTATTAAGTGAAGAAATGATCATGTGCTCCGAAGTCTGTGAGAGTCCATCTGCTGGA 2211  
 QY 621 ThrValLysHisGluGlyGlnIleLysTyrValAsnAspValArgLysIleThrLysLys 640  
 DB 2212 ACTGTTAAGCATGAAGCCAGATCAAAATATGCAATGACCGTCCGAAATCAACCAAGAAA 2271  
 QY 641 AsnIleGluGluTyrGlyProPheAspLeuValIleGlyGlySerProCysAsnAspLeu 660  
 DB 2272 AATATTGAAGAGTGGGCGCGTGTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2331  
 QY 661 SerAsnValAsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPhePheGlu 680  
 DB 2332 TCCTAAGCTCAATCTCTCCCAAGGTTATATGAGGGCAGAGAAAGCTCTCTCTTGGAG 2391  
 QY 681 PheTyrHisLeuLeuAsnTyrThrArgProLysGluGlyAspAsnArgProPhePheTyr 700  
 DB 2392 TTTTACCACTGTGTGATTTATACCCGCCCAAGAGAGGCGCAACCGTCAATCTTCTGG 2451  
 QY 701 MetPheGluAsnValValAlaMetLysValAsnAspLysLysAspIleSerArgPheLeu 720  
 DB 2452 ATGTTCGAGAAATGTTGGGCGCATGAAGTGAATGACAAAGAAACATCTCAAAATTCCTG 2511  
 QY 721 AlaCysAsnProValMetIleAspAlaIleLysValSerAlaAlaHisArgAlaArgTyr 740  
 DB 2512 GCATGTAAACCCAGTGTATGATGCAATCAAGGTGTCTGTCTCACAAGGGCGCGGTAC 2571  
 QY 741 PheTyrGlyAsnLeuProGlyMetAsnArgProValMetAlaSerLysAsnAspLysLeu 760



Db 2572 TTCTGGGTAACTACCCGAATGAAC----- 2598  
QY 761 GluLeuGlnAspCysLeuGluPheSerArgThrAlaLysLeuLysLysValGlnThrile 780  
Db 2598 ----- 2598  
QY 781 ThrThrLysSerAsnSerIleArgGlnGlyAsnGlnLeuPheProValValMetAsn 800  
Db 2598 ----- 2598  
QY 801 GlyLysAspValLeuTrpCysThrGluLeuGluArgLysPheGlyPheProAlaHis 820  
Db 2599 -----AGGATCTTCGGCTTCCCTGCTCAC 2622  
QY 821 TyrThrAspValSerAsnMetGlyArgGlyAlaArgGlnLysLeuLeuArgSerTrp 840  
Db 2623 TACACGGAGTGTCCACATGGCGCGCGCGCTCAGAGCTGCTGGGAGGCTTGG 2682  
QY 841 SerValProValIleArgHisLeuPheAlaProLeuLysAspTyrPheAlaCysGlu 859  
Db 2683 AGTGTACCGGTATCAGACACCTGTTTGCCTTGAAGGACTACTTTGCCTGTGAA 2739

RESULT 13  
LOCUS AF156488 4145 bp mRNA linear PRI 05-SEP-1999  
DEFINITION Homo sapiens DNA cytosine-5 methyltransferase 3 beta 1 (DNMT3B)  
ACCESSION AF156488  
VERSION AF156488.1 GI:5823167  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 4145)  
Xie,S., Wang,Z., Okano,M., Nogami,M., Li,Y., He,W.W., Okumura,K.  
and Li,E.  
Cloning, expression and chromosome locations of the human DNMT3  
gene family  
Gene 216 (1), 87-95 (1999)  
JOURNAL MEDLINE  
99365304  
PUBMED 10433969  
REFERENCE 2 (bases 1 to 4145)  
Xie,S. and Li,E.  
Direct Submission  
AUTHORS TITLE  
Submitted (03-JUN-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,  
Charlestown, MA 02129, USA  
JOURNAL Location/Qualifiers  
FEATURES  
source  
1. 4145  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/map="between D20S182 and D20S106; 39.9-50.2 cM"  
gene  
1. 4145  
/gene="DNMT3B"  
115..2676  
/gene="DNMT3B"  
/note="de novo DNA methyltransferase"  
/codon\_start=1  
/product="DNA cytosine-5 methyltransferase 3 beta 1"  
/protein\_id="AA053063.1"  
/translation="MKGDTIRHLNGEDAGREDSDILVNGACSDSDSPPILEAIRTP  
EIRGRSSRLSKREVSLSYTQDITGDDGDESDGSDTPVMPKLFRETRRSSEPA  
VTRNNNSVSRERHRPSPRSTRGRQNRHVDSPVEFPATRSRRRATASAGTPWPS  
PPSSVLTIDLTDDTDTGTPQSSPTPYARLAQDSQGGMESQVADSGDGSSEYQ  
DKGFGIGDLVNGIKGFSWPMVVMVSKATSKRQAMGMRVOWFGDGKFSVSADK  
LVALGFLSHENLATFNKLVSRKAMYHALEKARVRAGKTFPSSPGDSLQDLKPMLE  
WAHGGKPTGIEGLKPNPTOPVYKSKVRAGSKLESKYEKTRRTTADSNTPY  
CPAPKRLKTYNNGKDRGDEQSDREOMASDVANNKSSLEDGCLSCGRKNPVPFHP  
EGGCLQCDRDLFLFYDDGYSQYCTVCCGEBLLICSNFSCRCFCVCELVLP  
GTGAEEAKLQEPWSCVCLPQRCVGLVRRRDKDWNVLQAFFTSDTGLEVEAPKLYPA  
IPAAARRRPIRVLSFDGIATGYLVKLKELGKVGKYVASEVCEESIAVGTVKHEGNIKY

BASE COUNT 1041 a 1083 c 1096 g 925 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,26e-202 Length: 4145  
Score: 3759.00 Matches: 717  
Percent Similarity: 88.63% Conservative: 55  
Best Local Similarity: 82.32% Mismatches: 69  
Query Match: 81.90% Indels: 30  
DB: Gaps: 10  
US-09-720-086-6 (1-859) x AF156488 (1-4145)  
QY 1 MetLysGlyAspSerArgHisLeuAsnGluGluGluLysAlaSerGlyTyrGluGluCys 20  
Db 115 ATCAAGGGAGACACACAGGCATCTCAATGGAGAGGAGGACGCGCGGGAGGAGAACTCG 174  
QY 21 IleileValAsnGlyAsnPheSerAspGlnSerSerAspThrLysAspAlaProSerPro 40  
Db 175 ATCTCTGTCAACGGGGCTGTCAGGACGACGCTCTCCGAC-----TCGCC 219  
QY 41 ProValLeuGluAlaIleCysThrGluProValCysThrProGluThrArgGlyArgArg 60  
Db 220 CCAATCTGGAGGCTATC-----CCACACCCCGAGATCAGAGGCCGGAAGA 264  
QY 61 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuAsnTyrThrGlnAspMet 80  
Db 265 TCAAGCTCGGACTCTCCAAGAGGAGGTGTCAGTCTGCTAAGCTACACAGGACTTG 324  
QY 81 ThrGlyAspGlyAspArgAspGluValAspAspGlyAsnGlySerAspIle----Leu 99  
Db 325 ACAGGGCATGGCCACGGG-----GAAGATGGGGATGGCTCTGACACCCGACGTC 372  
QY 100 MetProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSerProAlaVal 119  
Db 373 ATGCCAAAGACTCTTCGGGAA-----ACCAGGACTCGTTTCAGAAAGCCAGCTGTC 423  
QY 120 ArgThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnArgAlaSerProAlaIle 139  
Db 424 CGAACTCGAAATAACACAGTGTCTCCAGCGGGAGAGGACACAGGCTTCCCGACGTTCC 483  
QY 140 ThrArgGlyArgGlnGlyArgHisHisValGlnGluTyrProValGluPheProAlaThr 159  
Db 484 ACCCGAGCGCGGAGGCGGCAACCATGTGGACGAGTCCCGCTGGAGTTCCCGGCTACC 543  
QY 160 ArgSerArgArgArgArgAlaSerSerSerAlaSerThrProTrpSerSerProAlaSer 179  
Db 544 AGTCTCTGAGCGGCGGCGCAACAGCATCGCAGAACCCATGGCCGCTCCCTCCCGAGC 603  
QY 180 -----ValAspPheMetGluGluVal-----ThrProLysSer 190  
Db 604 TCTTACCTTACCATCGACCTCACAGACGACACAGAGGACACACATGGGACGCCAGAGC 663  
QY 191 ValSerThrProSerValAspLeuSerGlnAspGlyAspGlnGluGlyMetAspThrThr 210  
Db 664 AGCAGTATCCCTTACGCCCCCTAGCCCGCCAGCAGCAGCAGGAGGAGGATGAGTCCCCG 723  
QY 211 GlnValAspAlaGluSerArgAspGlyAspSerThrGluTyrGlnAspAspLysGluPhe 230  
Db 724 CAGGTGGAGCAGACAGATGGAGATGGAGACAGATTCAGAGTATCAGGATGGGAGGAGTTT 783  
QY 231 GlyIleGlyAspLeuValTyrGlyLysIleLysGlyPheSerTrpTrpProAlaMetVal 250  
Db 784 GGAATAGGGGACCTCGTGTGGGGAAGATCAAGGGCTTCTCTGTGTGGCCCCCATGGTG 843  
QY 251 ValSerTrpLysAlaThrSerLysArgGlnAlaMetProGlyMetArgTrpValGlnTrp 270  
Db 844 GTGTCTTGAAGGCCACCTCCAAAGCGCAGAGCTATGTCTGGCATGCGGTGGGTCCAGTGG 903





```
/gene="DNMT3B"
322..2883
/gene="DNMT3B"
/codon_start=1
/product="DNA cytosine methyltransferase 3 beta"
/protein_id="AA157040.1"
/db_xref="GI:18033255"
/translation="MKGGTRHNGEEDAGGRDSDILVNGACSDQSDSDSPLEAHTTP
EIRRRSSRLSKREVSSLLSYDTLTDGGDGEDGSDTPVMPKLFRETRKSESPA
VTRNNNSVSRERHRSRSTRGRGRNHVDESPVEFPASTRLRRAATASAGTPWPS
PPSYLTIDLTDDTDTHTGTPQSSSTPYARLAQDSQGGMESPVQVEADSGDSSSYQ
DGKFGTIGDLVMGKI KGPSWMPAMVSWKATSKQAMSGMRVWFGDKGFSEVSADK
LVALGFSOHNELATFNKLVSRKAMHYALEKARVRAKTPFSSPDGSLDKLPKMLE
WANGGFPTGIEGLKPNNTQPVNKS KVRRAKSRKLESRYKTRRRRTADDASDTY
CPAPKRLKTCNYNGKRGDDEDSPEOMASOVANNKSSLEDGLSCGRKNPVSFHPLE
EGGLTAAKDRDFLELFYIMDDGYOSYCTVCCGRELILCSNTSCRCFCVCELEVLV
GTGTAAAEKLOEPWSCVLMCLPQRCRKHRRKRDNRVLRQAFRTSDTGLYEAPKLPA
IPAAARRPIRVLSFDGIATGVLVLEKIGIKVGVYVASEVCESTIAVGVKHEGNIKY
VNDVNTTKQNIEMGPFDPDLVIGSPDNLNVNPNARKGLYEGTGRLEFFHYHLLNYS
RPKEGDDRPFPFEMFENVAMKVGKRDISRFLCNPVMI DAIKVSAHRAVRVFWGNLP
GMNRPVIAKNDKLEQDCLYRNIKAKLVQTIITKNSNIKQGNQLFPVVMNGRED
VLWCTELERIFGFPVHYTVDVSNMGRGAKQLKLRGWSVPVIRHLFAPLKDYFACE"
BASE COUNT 1045 a 1172 c 1167 g 951 t
ORIGIN

Alignment Scores:
Pred. No.: 1.33e-202 Length: 4335
Score: 3759.00 Matches: 717
Percent Similarity: 88.63% Conservative: 55
Best Local Similarity: 82.32% Mismatches: 69
Query Match: 81.90% Indels: 30
DB: 9 Gaps: 10

US-09-720-086-6 (1-859) x AF331857 (1-4335)
QY 1 MetLysGlyAspSerArgHisLeuAenGluGluGluGlyAlaSerGlyTyrGluGluCys 20
Db 322 ATGAAGGAGACACACGAGCATCTCAATGGAGAGAGAGACCCCGCGGAGGAGAGATCG 381
QY 21 IleIleValAenGlyAsnPheSerAspGlnSerSerAspThrLysAspAlaProSerPro 40
Db 382 ATCTCGTCAACGGGGCCCTGCAGCGACCATCTCTCCGAC-----TCGCCC 426
QY 41 ProValLeuGluAlaIleCysThrGluProValCysThrProGluThrArgGlyArg 60
Db 427 CCAATCTGGAGGCTATC-----CGCACCCCGGAGATCAGAGGCGGAGA 471
QY 61 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuAenGlyThrGlnAspMet 80
Db 472 TCAACTCGCGACTCTCCAAGAGGGAGGTGTCCAGTCTGCTAAGCTACACACAGGACTTG 531
QY 81 ThrGlyAspGlyAspArgAspGluValAspAspGlyAsnGlySerAspIle---Leu 99
Db 532 ACAGGCGATGGCGACGG-----GAAGATGGGATGGCTCTGCACCCCGAGTC 579
QY 100 MetProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSerProAlaVal 119
Db 580 ATGCCAAAGCTCTTCGGGAA-----ACCAGGACTCGTTTCAGAAAGCCAGCTGTC 630
QY 120 ArgThrArgHisSerAenGlyThrSerSerLeuGluArgGlnArgAlaSerProArgIle 139
Db 631 CCAACTCGAATAACACACAGTGTCTCCAGCGGGAGAGGCACAGCCCTTCCCGAGTTC 690
QY 140 ThrArgGlyArgGlnGlyArgHisValGlnGluTyrProValGluPheProAlaThr 159
Db 691 ACCCGAGGCGCGGAGGCGGCAACCATGTGGACAGATCCCGCGGAGTCTCCCGCTACC 750
QY 160 ArgSerArgArgArgAlaSerSerSerAlaSerThrProTrpSerSerProAlaSer 179
Db 751 AGTCTCTCAGAGCGGCGGCAACAGCATCGGACGACGACGATGGCGCTCCCTCCAGC 810
QY 180 -----ValAspPheMetGluGluVal-----ThrProLysSer 190
Db 180 -----ValAspPheMetGluGluVal-----ThrProLysSer 190

/ gene="DNMT3B"
322..2883
/gene="DNMT3B"
/codon_start=1
/product="DNA cytosine methyltransferase 3 beta"
/protein_id="AA157040.1"
/db_xref="GI:18033255"
/translation="MKGGTRHNGEEDAGGRDSDILVNGACSDQSDSDSPLEAHTTP
EIRRRSSRLSKREVSSLLSYDTLTDGGDGEDGSDTPVMPKLFRETRKSESPA
VTRNNNSVSRERHRSRSTRGRGRNHVDESPVEFPASTRLRRAATASAGTPWPS
PPSYLTIDLTDDTDTHTGTPQSSSTPYARLAQDSQGGMESPVQVEADSGDSSSYQ
DGKFGTIGDLVMGKI KGPSWMPAMVSWKATSKQAMSGMRVWFGDKGFSEVSADK
LVALGFSOHNELATFNKLVSRKAMHYALEKARVRAKTPFSSPDGSLDKLPKMLE
WANGGFPTGIEGLKPNNTQPVNKS KVRRAKSRKLESRYKTRRRRTADDASDTY
CPAPKRLKTCNYNGKRGDDEDSPEOMASOVANNKSSLEDGLSCGRKNPVSFHPLE
EGGLTAAKDRDFLELFYIMDDGYOSYCTVCCGRELILCSNTSCRCFCVCELEVLV
GTGTAAAEKLOEPWSCVLMCLPQRCRKHRRKRDNRVLRQAFRTSDTGLYEAPKLPA
IPAAARRPIRVLSFDGIATGVLVLEKIGIKVGVYVASEVCESTIAVGVKHEGNIKY
VNDVNTTKQNIEMGPFDPDLVIGSPDNLNVNPNARKGLYEGTGRLEFFHYHLLNYS
RPKEGDDRPFPFEMFENVAMKVGKRDISRFLCNPVMI DAIKVSAHRAVRVFWGNLP
GMNRPVIAKNDKLEQDCLYRNIKAKLVQTIITKNSNIKQGNQLFPVVMNGRED
VLWCTELERIFGFPVHYTVDVSNMGRGAKQLKLRGWSVPVIRHLFAPLKDYFACE"
```

```
Db 811 TCTTACCTTACCATCGACTCACAGACGACACAGAGGACACACATGGAGCGCCCGAGAGC 870
QY 191 ValSerThrProSerValAspLeuSerGlnAspGlyAspGlnGluGlyMetAspThrThr 210
Db 871 AGCAGTACCCCTACCGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCCCTAGCC 930
QY 211 GlnValAspAlaGluSerArgAspGlyAspSerThrGluTyrGlnAspAspLysGluPhe 230
Db 931 CAGGTGGAGGACAGACAGTGGAGTGGAGACAGTTCAGAGTATCAGGATGGAGAGGATTT 990
QY 231 GlyIleGlyAspLeuValTrpGlyLysIleLysGlyPheSerTrpTrpProAlaMetVal 250
Db 991 GGAATAGGGGACCTCGTGTGGGGAAGATCAAGGGGCTTCTCTGGTGGCCCGCATGGTG 1050
QY 251 ValSerTrpLysAlaThrSerLysArgGlnAlaMetProGlyMetArgTrpValGlnTrp 270
Db 1051 GTGCTTGGAGGCCACCTCCAAAGCAGACAGGTATGTCTGCATGGCGTGGTCCAGTGG 1110
QY 271 PheGlyAspGlyLysPheSerGluIleSerAlaAspLysLeuValAlaLeuGlyLeuPhe 290
Db 1111 TTTGCGATGGCAAGTTCTCCGAGGTCTCTGCAGACAAACTGGTGGCAGCTGGGCTGTC 1170
QY 291 SerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLysAlaMetTyr 310
Db 1171 AGCCAGCACTTTAATTTGGCCACCTTCAATAGCTCCTCTCTATCGAAAGCCATGTAC 1230
QY 311 HisThrLeuGluLysAlaArgValArgAlaGlyLysThrPheSerSerProGlyGlu 330
Db 1231 CATGCTCGGAGGAGCTCAACCCCAACACACACACAGTGGTATTAATAGTCAAGTGGCT 1290
QY 331 SerLeuGluAspGlnLeuLysProMetLeuGluTrpAlaHisGlyGlyPheLysProThr 350
Db 1291 TCATTGGAGGACGAGCTGAAGCCCATGTTGGAGTGGGCCACCGGGGCTTCAAGCCCACT 1350
QY 351 GlyIleGluGlyLysProAsnLysLysGlnProValValAlaLysSerLysValArg 370
Db 1351 GGGATCGAGGCTCAACCCCAACACACACACAGTGGTATTAATAGTCAAGTGGCT 1410
QY 371 ArgSerAspSerArgAsnLeuGluProArgArgGluAsnLysSerArgArgThr 390
Db 1411 COTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1470
QY 391 ThrAsnAspSerAlaAlaSerGlu---SerProProLysArgLeuLysThrAsnSer 409
Db 1471 GCTGAGGACTCAGCCACCTCTGACTACTGCCCGCACCCCAAGCGCTCAAGACAAATTC 1530
QY 410 Tyr---GlyGlyLysAspArgGlyGluAspGluSerArgGluArgMetAlaSerGlu 428
Db 1531 TATAACAACGGCAACACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1590
QY 429 ValThrAsnAsnLysGlyAsnLeuGluAspArgCysLeuSerCysGlyLysAsnPro 448
Db 1591 GTTGCCCAACAACAGAGCAGCTGGAAGATGGCTGTTTGTCTTGTGGCAGGAAAAACCC 1650
QY 449 ValSerPheHisProLeuPheGluGlyLysGlnSerCysArgAspArgPheLeu 468
Db 1651 GTGCTCTTCCACCTCTCTTTGAGGGGGGCTCTGTGACACATGCCGGGATCGCTTCCTT 1710
QY 469 GluLeuPheTyrMetTyrAspGluAspGlyTyrGlnSerTyrCysThrValCysGlu 488
Db 1711 GAGCTCTTTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1770
QY 489 GlyArgGluLeuLeuLeuCysSerAsnThrSerCysCysArgCysPheCysValGluCys 508
Db 1771 GAGCGAGAGCTGTGCTTTTGCAGCAACACAGAGTGTGCTGCTGCTGCTGCTGCTGCTG 1830
QY 509 LeuGluValLeuValGlyAlaGlyThrAlaGluAspAlaLysLeuGlnGluProTrpSer 528
Db 1831 CTGAGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1890
QY 529 CysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgLysAspTrpAsn 548
Db 1891 TGTTACATGTGCTCTCCGCGAGCGCTGTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1950
```

QY 549 MetArgLeuGlnAspPheThrThrAspProAspLeuGluGlnPheGluProProlys 568  
Db 1951 GTGGCGCTGCAGAGCCCTTCTTCCACAGAGTACACAGGGCGCTT---GAATATGAGCCCCCAG 2007  
QY 569 LeuYrProAlaIleProAlaIleAlaAlaArgArgProIleArgValLeuSerLeuPheAsp 588  
Db 2008 CTGTACCTGCTCCATTCCTCCGAGCCCGAGCGCCCATTCAGTCTCTGATGTTTGAT 2067  
QY 589 GlyIleAlaThrGlyIleValLeuValLeuValLeuValLeuValLeuValLeuValLeu 608  
Db 2068 GGCACTCGGACAGGCTACCTAGTCTCTCAAGAGTTGGGCAATGAAGTGAAGAGTACGTC 2127  
QY 609 AlSerGluValCysAlaGluSerIleAlaValGlyThrValGlyHisGluGlyGlnIle 628  
Db 2128 GCTTCTGAGTGTGTAGAGAGTCCATGCTGTGGAACCGGAGAGCAGAGGGGAATATC 2187  
QY 29 LysTyrValAsnAspValArgLysIleThrLysLysAsnIleGluGluTyrGlyTyrPhe 648  
Db 2188 AATATCGTGAACGAGCTGAGAACATCAAAAGAAAAATATGAAGAAATGGGGCCCATTT 2247  
QY 649 AspLeuValIleGlyGlySerProCysAsnAspLeuSerAsnValAsnProAlaArgLys 668  
Db 2248 GACTTGGTATTGGCGAGGCGCATGCAACGATCTTCAATGTAATCCAGCCAGGAAA 2307  
QY 669 GlyLeuTyrGluGlyThrGlyArgLeuPhePheGluPheThrHisLeuLeuAsnTyrThr 688  
Db 2308 GGCTGTATGAGGGTACAGCGCGGCTCTTCTGAAATTTTACACACTGCTGAATTTACTCA 2367  
QY 689 ArgProLysGluGlyAspAsnArgProPheThrPhePheGluAsnValValAlaMet 708  
Db 2368 CGCCCCAGAGAGGCTATGACCGGCGCTCTTGTGAATGTTGAATTTGAGCCATG 2427  
QY 709 LysValAsnAspLysLysAspIleSerArgPheLeuAlaCysAsnProValMetIleAsp 728  
Db 2428 AAGGTTGGCGAAGAGGAGACATCTCACGGTCTGAGATGTAATCCATGATGATTTGAT 2487  
QY 729 AlaIleLysValSerIleAlaHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMet 748  
Db 2488 GCCATCAAAGTTCTGCTGCTCACAGGCGCGATCTTCTGGGCACTTACCCCGGATG 2547  
QY 749 AsnArgProValMetAlaSerLysAsnAspLysLeuGluLeuGlnAspCysLeuGluPhe 768  
Db 2548 AACGGCGCGTATGATCATCAAGATGATTAATCTGAGCTGCAGAGCTGCTTGGAATAC 2607  
QY 769 SerArgThrAlaLysLeuLysLysValGlnThrIleThrThrLysSerAsnSerIleArg 788  
Db 2608 AATAGGATAGCCAAAGTTAAAGAAAGTACAGACATTAACCAACCAAGTCCGAACTGATCAA 2667  
QY 789 GlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysAspAspValLeuTyrCys 808  
Db 2668 CAGGGGAAACCAACTTTTCCCTGCTGATGATGCAAAAGAGATGTTTGTGATGC 2727  
QY 809 ThrGluLeuGluArgIlePheGlyPheProAlaHisTyrThrAspValSerAsnMetGly 828  
Db 2728 ACTGAGCTCGAAGAGATCTTGGCTTCTGTCGACATCAACAGAGCTGCTCAACATGGGC 2787  
QY 829 ArgGlyAlaArgGlnLysLeuGluArgSerTyrPheValProValIleArgHisLeu 848  
Db 2788 CGTGTGCGCGCCAGAGAGCTGCTGGAGAGTCTCTGAGCGGCTGCTATCCACACACTC 2847  
QY 849 PheAlaProLeuLysAspTyrPheAlaCysGlu 859  
Db 2848 TTGCGCCCTCTGAAGGACTACTTTCATGTGA 2880  
RESULT 15  
AF176228 4267 bp mRNA linear PRI 26-OCT-1999  
LOCUS AF176228  
DEFINITION Homo sapiens DNA cytosine-5 methyltransferase 3B (DNMT3B) mRNA,  
complete cds.  
ACCESSION AF176228  
VERSION AF176228.1 GI:6118091  
KEYWORDS

SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE 1 (bases 1 to 4267)  
AUTHORS Xu,G.-L., Rescorl,T.H., Bourc'his,D., Hsieh,C.-L., Tommerup,N., Bugge,M., Hulten,M., Qu,X., Russo,J.V. and Viegas-Pequignot,E.  
TITLE Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene  
JOURNAL Nature (1999) In press  
REFERENCE 2 (bases 1 to 4267)  
AUTHORS Xu,G.L. and Bestor,T.H.  
TITLE Direct Submission  
JOURNAL Submitted (06-AUG-1999) Genetics and Development, Columbia University, 701 West 168 St., New York, NY 10032, USA  
FEATURES  
source  
1..4267  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/map="20q11-q13"  
1..4267  
/gene="DNMT3B"  
265..2802  
/note="mutated in human chromosome instability and immunodeficiency disease known as ICF syndrome; contains alternative 5' exon 1B"  
/codon\_start=1  
/product="DNA cytosine-5 methyltransferase 3B"  
/protein\_id="AA04015.1"  
/db\_xref="GI:6118092"  
/translation="MEPSPEPSSLESKMGDRHLNCEEDAGREDSILVNGACSDSS  
DSPLELAIPTPEIRGRSSRLSKREVSLSLYTODLTGGDGDGSDTPVMPKL  
FRFRTRESAPVATRRNNNSYSRRHRHSPRSRTSGQRNHHVDSPVPEPATRLSR  
RATASAGTWPSPSSYLITDLTDTEDEHTPQSSSTPYARLAODSOGGMSPOVE  
ADSGDSESEYODGKERGIDLVWGKIKGFSMPAMVMSRATSRQAGMSWVWF  
GDGKSEVSADKLVALGLFSQHPNLATFKLVGRKMYHLEKARVRAGKTPSSPG  
DSLBDQLKMLEAHAGSFKPTGIEGLKPNPTQENTRRRPADDSATSDYCPAPRLK  
TNCYNGKDRGDEDSREQMSADVANNKSLLEGCSLGRNPPVFPHPLEGGLCQT  
RDRFLLEPYMDDDGYOSYCTVCCGREGRELLCSNTSCRCCEVECLEVGTAAEA  
KLOEPMSCYMCIPORCHYRRKMDMYRLQAFSTPTGLEAEAPKLVPALPARRRP  
IRVSLPFGIATGYLVKELGDKVGVYASVEEBSLAVGYKHGNTKYNDYRNT  
KNLIEWGPFDLTVGSPCNDLSVNPAPKGLVETGRLPFEYHLLNLSRPEGDR  
PFPMFENVAMKVGDRIISRPLECNPMIDAIKVSAHAPRVFNGNIPGMNPVIA  
SKNDLKLBDQCLEYNRJAKLKKVOTITTSKNSIKQGNQLPVMWNGKEDVLWCTELE  
RIFGEPVHTDVSNMGRAROKLGRSWSVPIRHLFAPLKDYPACE"  
BASE COUNT 1069 a 1098 c 1147 g 953 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 7.01e-198 Length: 4267  
Score: 3675.00 Matches: 703  
Percent Similarity: 86.80% Conservative: 53  
Best Local Similarity: 80.71% Mismatches: 65  
Query Match: 80.07% Indels: 50  
Gaps: 11  
US-09-720-086-6 (1-859) x AF176228 (1-4267)  
QY 1 MetLysGlyAspSerArgHisLeuAsnGluGluGlyAlaSerGlyTyrGluGluCys 20  
Db 301 ATGAAAGGAGACACACGAGTCTCAATGAGAGAGAGACCGCCGCGAGGAAAGACTCG 360  
QY 21 IleIleValAsnGlyAsnPheSerAspGlnSerSerAspThrLysAspAlaProSerPro 40  
Db 361 ATCCTCGTCAACGGGCGCTGACGAGCAGTCTCCGAC-----TCGCCC 405  
QY 41 ProValLeuGluAlaIleCysThrGluProValCysThrProGluThrArgGlyArgArg 60  
Db 406 CCAATCTCGAGGCTATC-----CGACCCCGGAGATCAGAGCGCCAGAGA 450  
QY 61 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuAsnTyrThrIleAspMet 80

Db 451 TCAAGCTCGCCACCTCTCAAGAGGAGGTGTCCAGTCTGCTAAGCTACACACAGGACTTG 510  
Qy 81 ThrGlyAspGlyAspArgAspAspGluValAspAspGlyAsnGlySerAspIle---Leu 99  
Db 511 ACAGCGCATGGCGAGCGG-----GAAGATGGGATGGCTCTGCACACCCAGTC 558  
Qy 100 MetProLysLeuThrArgGluThrLysAspThrArgThrArgSerGluSerProAlaVal 119  
Db 559 ATGCCAAGCTCTCCGGAA-----ACCAGACTCGTTTCAGAAAGCCAGCTGTC 609  
Qy 120 ArgThrArgHisSerAsnGlyThrSerSerLeuGluArgGlnAlaSerProAlaGile 139  
Db 610 CGAACTCGAAATAACACAGTCTCTCCAGCGGAGAGGACACAGGCTTCCCCAGCTTC 669  
Qy 140 ThrArgGlyArgGlnGlyArgHisValGlnGlyProValGluPheProAlaThr 159  
Db 670 ACCGAGCGCGGAGCGGCGCAACCATGTGGAGAGTCCCGCTGGAGTTCCCGGCTACC 729  
Qy 160 ArgSerArgArgArgAlaSerSerAlaSerThrProTrpSerSerProAlaSer 179  
Db 730 AGGTCCCTGAGACGCGCGGCAACAGCATCGGACGAAACGCTATGGCGTCCCTCCAGC 789  
Qy 180 -----ValAspPheMetGluVal-----ThrProLysSer 190  
Db 790 TCTTACCTTACCATCGCTCTCAGACGACACACAGACACACATGGGACGCGCCAGAGC 849  
Qy 191 ValSerThrProSerValAspLeuSerGlnAspGlyAspGlnGluGlyMetAspThrThr 210  
Db 850 AGCAGTACCCCTACGCGCCCTAGCCAGGACAGCAGCGGCGGCACTGGATCCCGC 909  
Qy 211 GlnValAspAlaGluSerArgAspGlyAspSerThrGluTrpGlnAspAspLysGluPhe 230  
Db 910 CAGGTGGAGCAGACAGTGGAGTGCAGACAGTTCAGGTATCAGGATGGGAAGAGTTT 969  
Qy 231 GlyIleGlyAspLeuValTrpGlyLysIleLysGlyPheSerTrpTrpProAlaMetTyr 250  
Db 970 GGAATAGGGGACCTCGTGTGGGAAAGATCAAGGGCTCTCTGCTGGCGCCCATGGTG 1029  
Qy 251 ValSerTrpLysAlaThrSerLysArgGlnAlaMetProGlyMetArgTrpValGlnTrp 270  
Db 1030 GTGTCTTGAAGGCCACCTCCAAGCGACAGGCTATGTGCGCATGCGGTGGGTCCAGTGG 1089  
Qy 271 PheGlyAspGlyLysPheSerGluIleSerAlaAspLysLeuValAlaLeuGlyLeuPhe 290  
Db 1090 TTGGCGATGGCNAAGTCTCCGAGGTCTCTGCACACAACTGGTGGGCACCTGGGCTGTT 1149  
Qy 291 SerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLysAlaMetTyr 310  
Db 1150 AGCCAGCACTTTAATTTGGCCACCTCAATAAGCTCGTCTCTCTATCGAAGAGCCATGTAC 1209  
Qy 311 HisThrLeuGluLysAlaArgValArgAlaGlyLysThrPheSerSerSerProGlyGlu 330  
Db 1210 CATGCTCTGGAGAAAGCTAGGGTGGAGTGGAGCTGGCAAGACCTTCCCGAGCAGCGCTGGAGAC 1269  
Qy 331 SerLeuGluAspGlnLeuLysProMetLeuGluTrpAlaHisGlyGlyPheLysProThr 350  
Db 1270 TCATTGGAGACCGCTGAAGCCCATGTTGGAGTGGGCCCAACGCGGGCTTCAAGCCCACT 1329  
Qy 351 GlyIleGluGlyLeuLysProAsnLysGlnProValValAsnLysSerLysValArg 370  
Db 1330 GGGATCGAGCGCTCAAAACCAACACAGCAACCA----- 1365  
Qy 371 ArgSerAspSerArgAsnLeuGluProArgArgArgGluAsnLysSerArgArgArgThr 390  
Db 1366 -----GAGAACAAGACTCGAAGAGCGACA 1389  
Qy 391 ThrAsnAspSerAlaAlaSerGlu---SerProProLysLysArgLeuLysThrAsnSer 409  
Db 1390 GCTGACGACTCAGCCACTCTGACTACTGCCCCCGACCCAGCGCTCAAGACCAATTGC 1449  
Qy 410 Tyr---GlyGlyLysAspArgGlyGluAspGluGluSerArgGluArgMetAlaSerGlu 428  
|||

Db 1450 TATAACAACGCGCAAGACCGAGGGGATGAAGATCAGAGCGGAGAACAAATGGCTTCAGAT 1509  
Qy 429 ValThrAsnAsnLysGlyAsnLeuGluAspArgCysLeuSerCysGlyLysLysAsnPro 448  
Db 1510 GTTGCAACAACAAGAGCAGCCTCGAAGATGGCTGTGTTTCTTCTGTGCGCAGGAAAAACCCC 1569  
Qy 449 ValSerPheHisProLeuPheGluGlyGlyLeuCysGlnSerCysArgAspArgPheLeu 468  
Db 1570 GTGTCCTTCCACCTCTCTTTGAGGGGGGGCTCTGTACACATGCGCGGATCGCTTCCT 1629  
Qy 469 GluLeuPheTyrMetTyrAspGluAspGlyTyrGlnSerTyrCysThrValCysCysGlu 488  
Db 1630 GAGCTGTTTTATCATGTATGATGACGATGGCTATCAGCTTCTTACTGTCACGTGTGTGCGAG 1689  
Qy 489 GlyArgGluLeuLeuLeuCysSerAsnThrSerCysCysArgCysPheCysValGluCys 508  
Db 1690 GGCCGAGAGCTGCTCTCTTTCAGCAACACAGCTGCTGCGGTGTTCTCTGTGTGGAGTGC 1749  
Qy 509 LeuGluValLeuValGlyAlaGlyThrAlaGluAspAlaLysLeuGlnGluProTrpSer 528  
Db 1750 CTGGAGGTGCTGTGTGGGCACAGGCACAGCGCGCCGACGCTTCAGAGAGCCCTGGAGC 1809  
Qy 529 CysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgArgLysAspTrpAsn 548  
Db 1810 TGTTCATGTGTCTCCGAGCGCTGTCTATGCGCTCTGCGGCGCGGAAAGGACTGGAAC 1869  
Qy 549 MetArgLeuGlnAspPheThrThrAspProAspLeuGluGluPheGluProLys 568  
Db 1870 GTGCGCTGACGGCTTCTTCCAGCTGACACGGGCTT---GAATATGAAGCCCCCAAG 1926  
Qy 569 LeuTyrProAlaIleProAlaLysArgArgProIleArgValLeuSerLeuPheAsp 588  
Db 1927 CTGTACCTTCCCTATCCCGCAGCCGAGCGGCCCATTCAGTCTCTCTCATGTTGTAT 1986  
Qy 589 GlyIleAlaThrGlyTyrLeuValLeuLysGluLeuGlyIleLysValGluLysTyrIle 608  
Db 1987 GGCACTCGCAGCAGGCTACCTAGTCTCTCAAGAGTGTGGGCATAAAGTAGGAAGTACGTC 2046  
Qy 609 AlaSerGluValCysAlaGluSerIleAlaValGlyThrValLysHisGluGlyGlnIle 628  
Db 2047 GCTTCTGAAGTGTGTGAGGATCCATGCTGTGGAACCGTGAAGCAGCAGGGAATATC 2106  
Qy 629 LysTyrValAsnAspValArgLysIleThrLysLysAsnIleGluGluTrpGlyProPhe 648  
Db 2107 AAATAGCTGAACGACGTGAGGAACATCACAAAGAAATAATTTGAAGAAATGGGGCCATTT 2166  
Qy 649 AspLeuValIleGlySerProCysAsnAspLeuSerAsnValAsnProAlaArgLys 668  
Db 2167 GACTTGGTGTATGGCGGAAGCCCATGCAACGATCTCTCAATGTGAATCCAGCCAGGAAA 2226  
Qy 669 GlyLeuTyrGluGlyThrGlyArgLeuPhePheGluPheTyrHisLeuLeuAsnTyrThr 688  
Db 2227 GGCTGTATGAGGTACAGCCGGCTCTTCTTCAATTTTACCACCTGCTGTAATTACTCA 2286  
Qy 689 ArgProLysGluGlyAspAsnArgProPheTrpMetPheGluAsnValAlaMet 708  
Db 2287 CGCCCCAAGAGGGTGTATGACCGCGCTTCTTCTGGATGTTTGAGAAATGTTGTAGCCATG 2346  
Qy 709 LysValAsnAspLysLysAspIleSerArgPheLeuAlaCysAsnProValMetIleAsp 728  
Db 2347 ARGTTGGCGACAAAGGACATCTACGCTTCTGAGTCTTAATCCAGTGTATGTAT 2406  
Qy 729 AlaIleLysValSerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMet 748  
Db 2407 GCCATAAAGTTTCTGTCTGTCTACAGGGCCGATCTTCTGGGGCACTTACCCCGGATG 2466  
Qy 749 AsnArgProValMetAlaSerLysAsnAspLysLeuGluLeuGlnAspCysLeuGluPhe 768  
Db 2467 AACAGGCCCTGTAGCATCAAGAATGATAAATCGAGCTCGAGGACTGCTTGAATATC 2526  
Qy 769 SerArgThrAlaLysLeuLysValGlnThrIleThrThrLysSerAsnSerIleArg 788  
Db 2527 AATAGGATGCCAAGTTAAGAAAGTACAGACAAATACCAACCAAGTCAACTCGATCAAA 2586  
|||

```

OY 789 GlnGlyLysAsnGlnLeuPheProValMetAsnGlyLysAspAspValLeuTrpCys 808
    |||||
Db 2587 CAGGGGAAACCACTTTTCCCTGTGTGATGAATGGCAAGAAGATGTTTGTGTGTC 2646
    |||||
OY 809 ThrGluLeuGluArgIlePheGlyPheProAlaHisTyrThrAspValSerAsnMetGly 828
    |||||
Db 2647 ACTGAGCTCGAAAGGATCTTGGCTTTCCTGTGCACTACACAGAGTGTCCAAACATGGGC 2706
    |||||
OY 829 ArgGlyAlaArgGlnLysLeuGlyArgSerTrpSerValProValIleArgHisLeu 848
    |||||
Db 2707 CGTGTGCCCCGCCAGAGCTGTGTGGAAAGTCTCTGGAGCGTGTCTCATCCGACACCTTC 2766
    |||||
OY 849 PheAlaProLeuLysAspTyrPheAlaCysGlu 859
    |||||
Db 2767 TTGCCCCCTCTTGAAGGACTACTTGTGATGTGAA 2799
    |||||

```

Search completed: November 22, 2002, 10:52:58  
 Job time : 4989.5 secs

**THIS PAGE BLANK (USPTO)**

sequence 301 BP; 96 A; 65 C; 63 G; 66 T; 11 other;

Matches 241; Conservative 0; Mismatches 57; Indels 6; Gaps 3

Db 298 TAAA 301

CC determined (esp. using primers and probes derived from the GS

**THIS PAGE BLANK (USP)**



AF156488  
LOCUS AF156488 4145 bp mRNA linear PRI 05-SEP-1999  
DEFINITION Homo sapiens DNA cytosine-5 methyltransferase 3 beta 1 (DNMT3B)  
ACCESSION AF156488.1 GI:5823167  
VERSION AF156488.1  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS 1 (bases 1 to 4145)  
Xie,S., Wang,Z., Okano,M., Nogami,M., Li,Y., He,W.W., Okumura,K.  
TITLE Cloning, expression and chromosome locations of the human DNMT3 gene family  
JOURNAL Gene 236 (1), 87-95 (1999)  
MED 99365304  
10433969  
2 (bases 1 to 4145)  
Xie,S. and Li,E.  
REFERENCE Direct Submission  
Submitted (03-JUN-1999) CVRC, Mass. Gen. Hospital, 149 13th Street, Charlestown, MA 02129, USA  
JOURNAL  
TITLE  
AUTHORS  
REMARKS  
FEATURES  
source  
1. 4145  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/map="between D20S182 and D20S106; 39.9-50.2 cM"  
1. 4145  
/gene="DNMT3B"  
115. 2676  
/gene="DNMT3B"  
/note="de novo DNA methyltransferase"  
/product="DNA cytosine-5 methyltransferase 3 beta 1"  
/codon\_start=1  
/protein\_id="A05306.1"  
/db\_xref="GI:5823167"  
/translation="MKGDTIRHLNGBEDAGREDISLVNACSDSDSPPILEAIRP  
EIRGRSSRLSKREVSSLSYTQDLTGCGDGDSDPTVMPPLFRERTSESA  
VTRNNNSVSRHRSPRSTRGRQGNHVDSPVEPFAIRSLRRAATASATPMP  
PPSYLTITDITDTEDEHGTQSSSTPYARLADSOQCGMESPOVEADSGDSBYO  
DKKEFGIDLVNGKIKGFSWPAWVSWKATSKRAMSGMWQVGDGKFSVADK  
LVALGLFSQHPNLTATFNKLVSYRKAMTHALEKRAVAGKTPSPSDLEDOLKPMLE  
WAGGFKPFGIEGLKPNNTQPVNKSIVRAGSRKLSRKTEKTRRTAADSATSDY  
CAPARKTCYNNKXGDRGDESDREOMASDVANNSLEDCLSGCRKAPVSTHPLF  
EGGLCOTCDRLLEFYVDDGYSYCTVCGRELLCGSTSCRCFCEVCELVY  
GTGTAEAALQEPWSCYVCLPDRCHSVLRKXDMNVRLOAFPTSDGLEAKHEGNIKY  
IPARRRPRLVLSLFDGATGYLVLEKELIKVGVASVEVCESTIAGTAKHEGNIKY  
VNDVNTTKNIEEMGPDLYIGSPCNDISNVPARGLYEGTGLFFEFYHLNYS  
RPEGDDRRPFPMENYAMKVGDKDISFLFCNPVNIIDAKVSAHARARFYFWGKLP  
GNNRPVTSKNDKLELDCLEENRIAKLKKVOTITTKSNSIKQGNOLPVMVNGKED  
VLMCTELERIFGPFVHYTDSNNGRAGKQLKGRSWSVPIRHLFAPLKDYFACE"

gene  
CDS  
BASE COUNT 1041 a 1083 c 1096 g 925 t  
ORIGIN

Alignment Scores:  
Pred. No.: 5.02e-234 Length: 4145  
Score: 4566.00 Matches: 853  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 100.00% Gaps: 0  
DB: 9

US-09-720-086-8 (1-853) x AF156488 (1-4145)

QY 1 MetLVAGIYAspThrArgHisLeuAenGlyLgluAaspAlGlyGlyArgGluAspSer 20  
DB 115 ATGAAGGAGACACCGCATCTCAATGACAGAGAGCGCGCGGAGGAGAACTCG 174  
21 TleuValaenGlyAlaCysSerAaspGlnSerSerAaspSerProProlleuGluAla 40

DB 175 ATCTCTGTCACAGGGGCGCTGACAGACCAAGTCTCCGACTCGCCCCCAATCTTGAGAGCT 214  
QY 41 TleArgThrProGluIleArgGlyArgArgSerSerArgLeuSerLysArgGluVal 60  
DB 235 ATCCGACACCCCGAGATCAGAGCCGGAAGACCAAGCTCGACTCTCCAGAGAGAGCTG 294  
QY 61 SerSerLeuLeuSerTyrThrGlnAspLeuThrGlyAaspGlyAaspGlyAasp 80  
DB 295 TCCAGTCTGTAAAGCTACACACAGAGACTTGAACAGGATGCGACCGGGAAGATGGGAT 354  
QY 81 GlySerAspThrProValMetProLysLeuPheArgGluThrArgThrArgSerGluSer 100  
DB 355 GGCTGTGACACCCCGACTGATGCGCAAGCTCTCCGGGAACCGAGACTCTGTCAGAAAGC 414  
QY 101 ProAlaValArgThrArgAsnAenAsnSerValSerSerArgGluArgHisArgProSer 120  
DB 415 CCAGCTGTCCGAATCGAATAAACAAGTGTCTCCAGCCGGGAGAGCACAGGCTTCC 474  
QY 121 ProArgSerThrArgGlyArgGlnGlyArgAsnHisValAspGluSerProValGluPhe 140  
DB 475 CCAGCTTCCACCCCGAGGCGCGGACGCGGACCATGTGACAGACTCCCGCTGAGATTCC 534  
QY 141 ProAlaThrArgSerLeuArgArgArgAlaThrAlaSerAlaGlyThrProTyrProSer 160  
DB 535 CCGGTACACAGGTCCTTGACAGCGCGGCGGACAGCATCGGACGAGAACGCTGCGCTCC 594  
QY 161 ProProSerSerTyrLeuThrIleAspLeuThrAspAspThrGluAspThrHisGlyThr 180  
DB 595 CCTCCAGCTCTTACCTTACATCACTCAGACAGACAGACAGACAGACAGACAGACAG 654  
QY 181 ProGlnSerSerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyGlyMet 200  
DB 655 CCCAGAGACGAGTACCCCTTACCGCCGCTGAGCCAGACAGACAGAGGAGGAGCATG 714  
QY 201 GlnSerProGlnValGluAlaAspSerGlyAaspGlyAaspSerSerGlyTyrGluAaspGly 220  
DB 715 GAGTCCCGGAGGTGAGGAGGACAGCATGAGATGAGACAGTTCACAGTATCAGAGATGAG 774  
QY 221 LysGluPheGlyIleGlyAaspLeuValTyrGlyLysIleGlyGlyPheSerTyrTyrPro 240  
DB 775 AAGGAGTTTGAAATAGGGGACCTCGTGTGGGAAAGATCAAGGCTTCTCTGTGGGCC 834  
QY 241 AlaMetValValSerTyrLysAlaThrSerLysArgGlnAlaMetSerGlyMetArgTyr 260  
DB 835 GCCATGGTGTGTCTTGGAAAGCCACCTCCAAAGACAGACAGCTGTGCGATCGGTG 894  
QY 261 ValGlnTyrPheGlyAaspGlyLysPheSerGluValSerAlaAspLysLeuValAlaLeu 280  
DB 895 GTCCAGTGTGTGCGATGCGCAAGTCTCCAGAGTCTCCAGACAAACAGTGTGCGACTG 954  
QY 281 GlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLys 300  
DB 955 GGGCTGTTCAGCAGCAGCACTTAATTGGCCACCTCAATTAAGCTCGTCTCCATCGAAAA 1014  
QY 301 AlaMetTyrHisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSerSer 320  
DB 1015 GCCATGTACCAATGCTCGGAGAAAGCTAGGTCGAGCTGCGCAAGACCTTCCCGACAGC 1074  
QY 321 ProGlyAaspSerLeuGluAaspGlnLeuLysProMetLeuGluTyrPheAlaHisGlyGlyPhe 340  
DB 1075 CCGTGAAGCTCATTTGAGAGACACAGCTGAAGCCCATGTTGGAGTGGCCACCGGGGGCTTC 1134  
QY 341 LysProThrGlyIleGlyGlyLysLeuLysProAsnAsnThrGlnProValAlaAsnLysSer 360  
DB 1135 AAGCCACCTGGAGTCAGGGGCTCAACCAACCAACGACGCAACGCTGTTAATTAAGTCG 1194  
QY 361 LysValArgArgAlaGlySerArgLysLeuGlnSerArgLysTyrGluAsnLysThrArg 380  
DB 1195 AAGGTGCGTGTGTCAGCAGCAGTGAAGAAATTAAGATCAAGGAATAGAGAAACAGACTCG 1254  
QY 381 ArgArgThrAlaAspAspSerAlaThrSerAspTyrCysProAlaProLysArgLeuLys 400  
DB 1255 AGACGACAGCTGACGACTGACGACCTGACTACTGCCCCGACCAAGCGGCTCAAG 1314

**THIS PAGE BLANK (USPTO)**

OY 401 ThrAenCYTyrAenbnglYlYsAspArgGlyAAspGluAAspGlnSerArgGluGlnMet 420  
DB 1315 ACAATTGCTATACAAACGGCAAGACCGAGGGATGAGATCAGAGCCGAGAACAAATG 1374  
OY 421 AlaSerAAspValAlaAsnAsnLysSerSerLeuGluAAspGlyCYsLeuSerCYsGlyArg 440  
DB 1375 GCTTCAGATGTTGCCAACACAGACAGCTGGAAAGATGGCTGTTTCTTGTGGCAGG 1434  
OY 441 LysAAspProValSerPheHisProLeuPheGluGlyLeuCYsGlnThrCYsArgAsp 460  
DB 1435 AAAAACCCTGCTCTTCCACCTCTTCTTGAAGGGGGGCTCTGTACAGACATGCCGGGAT 1494  
OY 461 ArgPheLeuGluLeuPheTYrMetTYrAspAspAspGlyTYrGlnSerTYrCYsThrVal 480  
DB 1495 CGCTTCCTTGACCTGTTTACATGATGATGAGATGGCTATCAAGCTTACTGCACCTGTG 1554  
OY 481 CYsCYsGluGlyArgGluLeuLeuLeuCYsSerAsnThrSerCYsArgCYsPheCYs 500  
DB 1555 TGCTGCGAGGGCGAGAGCTGCTGTTTGACCAACAGACAGCTGCTGCCGTGTTCTGT 1614  
OY ValGluCYsLeuGluValLeuValGlyThrGlyThraAlaGluAlaLysLeuGlnGlu 520  
DB GTGAGTGCCTGAGAGGTGCTGCTGGCACAGCACAGCGCCGAGCCCAAGCTTCAGAG 1674  
OY 521 ProTrpSerCYsTYrMetCYsLeuProGlnArgCYsHisGlyValLeuArgArgLys 540  
DB 1675 CCTCGAGAGCTGCTACATGCTGCTCCGACGCTGCTCATGCGCTGCGCGCGGAG 1734  
OY 541 AspTrpAsnValArgLeuGlnAlaPhePheThrSerAspThrGlyLeuGluTYrGluAla 560  
DB 1735 GACTGGAACGTGGCGCTGCGACGGCTTCTTCCACAGTGACACAGGCTTGAATACGAAGCC 1794  
OY 561 ProLysLeuTYrProAlaAlaProAlaAlaArgArgProIleArgValLeuSerLeu 580  
DB 1795 CCCAAGCTGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1854  
OY 581 PheAspGlyIleAlaThrGlyTYrLeuValLeuLysGluLeuGlyIleLysValGlyLys 600  
DB 1855 TTGATGCGATCGCAGACAGGCTACCTAGCTCAAGAGTTGGCATTAAGTAGGAAG 1914  
OY 601 TYrValAlaSerGluValCYsGluGlnSerIleAlaValGlyThraLysHisGluGly 620  
DB 1915 TACGTCGCTTCTGAAGTGTGAGAGATGCTCATGCTGTTGGAACCTGGAAGCAGAGGG 1974  
OY 621 AsnIleLysTYrValAsnAspValArgAsnIleThrLysLysAsnIleGluGluTrpGly 640  
DB 1975 AATATCAATACCTGAACGAGCTGAGGAATCATCAAAAGAAATATTGAAGATGGGGC 2034  
OY 641 ProPheAspLeuValIleGlyGlySerProCYsAsnAspLeuSerAsnValAsnProAla 660  
DB CCATTGACTTGTGATGGCGGAAGCCCATGCAACGATCTCTCAATGTGAATCCAGCC 2094  
OY 661 ArgLysGlyLeuTYrGluGlyThrGlyArgLeuPhePheGluPheTYrHisLeuLeuAsn 680  
DB 2095 AGGAAGGCTGTATGAGGGTACAGGCCGCTCTTCTGCAATTTACCACTGCTGAAT 2154  
OY 681 TYrSerArgProLysGluGlyAspAspArgProPhePheTrpMetPheGluAsnValVal 700  
DB 2155 TACTCAGCGCCCAAGAGGGGTATGACCGGCTCTTCTGATGTTTGAAGATGTGTA 2214  
OY 701 AlaMetLysValGlyAspLysArgAspIleSerArgPheLeuGluCYsAsnProValMet 720  
DB 2215 GCCATGAAGGTGGCGACMAAGGGACATCTCAAGGTTCTGAGAGTAACTCAAGTATG 2274  
OY 721 IleAspAlaIleLysValSerAlaAlaHisArgAlaArgTYrPheTrpGlyAsnLeuPro 740  
DB 2275 ATGATGCGATCAAAAGTTTCTGCTGCTCAAGGGCCGATACTTCTGGGGCAACCTTACC 2334  
OY 741 GlyMetAsnArgProValIleAlaSerLysAsnAspLysLeuGluLeuGlnAspCYsLeu 760  
DB 2335 GGGATGAACAGGCGCGTATGATCAAAAGATGATAACTGAGCTGCAGACTGCTTG 2394

OY 761 GluTYrAsnArgIleAlaLysLeuLysLysValGlnThrIleThrThrLysSerAsnSer 780  
DB 2395 GAATACATAGATAGCCAAAGTTAAAGAAAGTACAGACAAATTAACCAAGTGAAGCTCG 2454  
OY 781 IleLysGlnGlyLysAsnGlnLeuPheProValIleMetAsnGlyLysGluAspValLeu 800  
DB 2455 ATCAACAGGGGAAAAACCACTTTCCCTGTTGTCATGAATGCAAGAGATGTTTG 2514  
OY 801 TrpCYsThrGluLeuGluArgIlePheGlyPheProValHisTYrThrAspValSerAsn 820  
DB 2515 TGCTGACTAGACTGAAAGATCTTGGCTTCTGTCGACTACACAGACGCTGCCAAC 2574  
OY 821 MetGlyArgGlyAlaArgGlnLysLeuLeuGlyArgSerTrpSerValProValIleArg 840  
DB 2575 ATGGGCGGTGTGCTGCCGACGAAAGCTGCTGGAAGGCTCTGAGCGTGCATCGGA 2634  
OY 841 HisLeuPheAlaProLeuLysAspTYrPheAlaCYsGlu 853  
DB 2635 CACCTTTCGCCCTCTGGAAGACTACTTGCATGTCAA 2673

RESULT 2  
AF331857  
LOCUS  
DEFINITION Name: AF331857

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 22, 2002, 03:58:32 ; Search time 4860.31 Seconds  
(without alignments)  
5107.633 Million cell updates/sec

Title: US-09-720-086-8  
Perfect score: 4566  
Sequence: 1 MKGDRHLNGEBDAGGRDS.....MSVPVIRHLFAPLKDYFACE 853

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=xlp  
-Q=/cogn2.1/USPRO.spool/US09720086/runat.18112002.092832.22195/app.query.fasta.1.4252  
-DB=GenBdb1 -QMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=prc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09720086@cogn2.1.11308@runac.18112002.092832.22195 -NCPU=6 -ICPU=3  
-NO\_XLPHY -NO\_MAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOB=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenBdb1.\*  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_in:\*  
18: em\_mu:\*  
19: em\_om:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_un:\*  
28: em\_un:\*

29: em\_vi:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_inv:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rod:\*  
36: em\_htg\_mam:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4566	100.0	4145	9 AF156488	AF156488 Homo sapi
2	4566	100.0	4335	9 AF331857	AF331857 Homo sapi
3	4450	97.5	4267	9 AF176228	AF176228 Homo sapi
4	4083.5	89.4	3897	9 AF156487	AF156487 Homo sapi
5	3759	82.3	4195	10 AF068626	AF068626 Mus muscu
6	3748	82.1	4223	10 AF151973	AF151973 Mus muscu
7	3748	82.1	4338	10 AF151969	AF151969 Mus muscu
8	3675	80.5	4135	10 AF068627	AF068627 Mus muscu
9	3664	80.2	4163	10 AF151974	AF151974 Mus muscu
10	3664	80.2	4278	10 AF151970	AF151970 Mus muscu
11	3415.5	74.6	4006	10 AY078427	AY078427 Mus muscu
12	3404.5	74.6	4034	10 AF151975	AF151975 Mus muscu
13	3404.5	74.6	4149	10 AF151972	AF151972 Mus muscu
14	3331.5	73.0	3946	10 AF068628	AF068628 Mus muscu
15	3320.5	72.7	3974	10 AF151976	AF151976 Mus muscu
16	3320.5	72.7	4089	10 AF151971	AF151971 Mus muscu
17	2767.5	60.6	3017	9 AK001191	AK001191 Homo sapi
18	2767.5	60.6	4094	10 BC007466	BC007466 Mus muscu
19	2170.5	47.5	3005	9 AF067972	AF067972 Homo sapi
20	2170.5	47.5	4192	10 AF068625	AF068625 Mus muscu
21	2170.5	47.5	4258	9 AF331856	AF331856 Homo sapi
22	1834.5	40.1	2948	5 AF135438	AF135438 Dario rer
23	1740	38.1	2191	9 AK025230	AK025230 Homo sapi
24	1414	31.0	176697	2 AC112586	AC112586 Rattus no
25	1313	28.8	123936	2 AC120824	AC120824 Rattus no
26	1240	27.2	1758	9 BC018214	BC018214 Homo sapi
27	1183.5	25.9	167568	2 AC111734	AC111734 Rattus no
28	1178	25.8	2127	9 AF129267	AF129267 Homo sapi
29	1161.5	25.4	168651	9 AC009474	AC009474 Homo sapi
30	1020.5	22.3	2057	9 AF129268	AF129268 Homo sapi
31	926	20.3	2008	9 AF129269	AF129269 Homo sapi
32	903	19.8	2077	6 AR129189	AR129189 Sequence
33	889.5	19.5	244329	2 AC107644	AC107644 Mus muscu
34	806.5	17.7	110000	2 AL354832_0	AL354832 Homo sapi
35	806.5	17.7	118899	9 HS1085F17	AL035071 Human DNA
36	751	16.4	225045	2 AL833803	AL833803 Mus muscu
37	751	16.4	249245	2 AC122356	AC122356 Mus muscu
38	671.5	14.7	1397	9 BC002560	BC002560 Homo sapi
39	669	14.7	1705	9 AF194032	AF194032 Homo sapi
40	661.5	14.5	1589	10 AF220524	AF220524 Mus muscu
41	661.5	14.5	1645	10 MM040467	AJ404467 Mus muscu
42	646	14.1	188936	2 AC128509	AC128509 Rattus no
43	615	13.5	210269	2 AC116459	AC116459 Mus muscu
44	608.5	13.3	119630	2 AC112040	AC112040 Rattus no
45	520	11.4	153533	2 AC095199	AC095199 Rattus no

RESULT 1

ALIGNMENTS

AF156488 4145 bp mRNA linear PRI 05-SEP-1999  
 LOCUS Hmo sapiens DNA cytosine-5 methyltransferase 3 beta 1 (DNMT3B)  
 DEFINITION mRNA, complete cds.  
 ACCESSION AF156488  
 VERSION AF156488.1 GI:5823167  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 4145)  
 Xie, S., Wang, Z., Okano, M., Li, Y., He, W.W., Okumura, K.  
 and Li, E.  
 Cloning, expression and chromosome locations of the human DNMT3  
 gene family  
 Gene 236 (1), 87-95 (1999)  
 JOURNAL MEDLINE  
 PUBMED 99365304  
 10433969  
 REFERENCE 2 (bases 1 to 4145)  
 Xie, S. and Li, E.  
 Direct Submission  
 TITLE Submitted (03-JUN-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,  
 Charlestown, MA 02129, USA  
 JOURNAL Location/Qualifiers  
 FEATURES  
 source  
 1..4145  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="20"  
 /map="between D20S182 and D20S106; 39.9-50.2 cm"  
 1..4145  
 /gene="DNMT3B"  
 115..2676  
 /gene="DNMT3B"  
 /note="de novo DNA methyltransferase"  
 /product="DNA cytosine-5 methyltransferase 3 beta 1"  
 /codon\_start=1  
 /protein\_id="AAD53063.1"  
 /translation="MKGDTRLHNGEEDAGGREDISLVNGACSDQSDSPPIRLAIRTP  
 EIRGRSSRLSKREVSLVSYTDITDGDGDDGSDTPVMPKLFRETRRSESPPA  
 VTRNNSSVSRHRPSPSTRGROGRNVHDESVEFFPATRLRRRATASAGTPWS  
 PSSVLTIDLTDDTHGTPOSSPTPYARLAQDSQGGMESQVADSDGDSSEYQ  
 DGKRGIGLVKGIKGFSSWPMVYVSKATSKRAMSGMRVQWFGDGFVSADK  
 LVALGFLSQHENLATFNKLVYRKAMVHALKXARVAGKTFPSRPGDSLDDLKPMLE  
 WAHGFKPTIGLELPNTQPVYVNSKVRAGSRKLESKYENKTRPRTRADDGNTSDY  
 CPAPKRLTCYNNNGRDEDOSEQMASDVANNKSLSDGGLSCGRKNPVSPHPLF  
 EGGLCQTCRDLFLFYDDGYQSYCTVCCGRELILLCNSCCRCFCVCELELV  
 GTGTAAEAKLQEPWSCYMCIPQRCVGRVRRKDMVRLQAFSTDGLVEAPKLYPA  
 IPAARRPRLVLSLFDGIATGVLVLELGT KVGYVASEVCEESIAVGTVKEGNIKY  
 VNDVNTTKNTIEEWGPFDLVIGGSPCNOLSNVNPARKGLYEGTGLFFEFYHLNYS  
 RPKEGDRPFFWFMENVMKVGDKRDIGRFLFCNPVMDAIKVSAARHARYFWGNLP  
 GMRPVIASKNDKLELDQCLEVNRVIAKLKVOTITTKNSNISKQGNOLFPVWNGKED  
 VLMCTELERIPGFPHVHTVDSNMGRAROKLLGRSNVPIRHLFAPLKYFACE"  
 BASE COUNT 1041 a 1083 c 1096 g 925 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 5,02e-234 Length: 4145  
 Score: 4566.00 Matches: 853  
 Percent Similarity: 100.00 Conservative: 0  
 Best Local Similarity: 100.00 Mismatches: 0  
 Query Match: 100.00 Indels: 0  
 DB: 9 Gaps: 0

US-09-720-086-8 (1-853) x AF156488 (1-4145)

Qy 1 MetLysGlyAspThrArgHisLeuAsnGlyGluGluAspAlaGlyArgGluAspSer 20  
 Db 115 ATGAAGGGAGACACCGGCAATCTCAATGGAGAGAGGACCGCGGGGAGGAACTCG 174  
 Qy 21 IleLeuValAsnGlyAlaCysSerAspGlnSerAspSerProIleLeuGluAla 40

Db 175 ATCTCTGTCACAGGGGCTGACGACGACAGTCTCTCGACTCGCCCCCAATCTCTGGAGGCT 234  
 Qy 41 IleArgThrProGluIleArgGlyArgArgSerSerArgLeuSerLysArgGluVal 60  
 Db 235 ATCCGACCCCGGAGATCAGAGGCGGAGATCAAGATCGCGACTCTCCAAGAGGAGGTG 294  
 Qy 61 SerSerLeuLeuSerTyrThrGlnAspLeuThrGlyAspGlyAspGlyGluAspGlyAsp 80  
 Db 295 TCCAGTCTGCTAAGCTACACACAGGACTTGACAGGCGATGCGCGGGGAAGATGGGAT 354  
 Qy 81 GlySerAspThrProValMetProLysLeuPheArgGluThrArgThrArgSerGluSer 100  
 Db 355 GGCTCTGACACCCCGAGTCATGCCAAAGCTCTTCGCGGAAACACGAGGACTCTCTTCAAGAAGC 414  
 Qy 101 ProAlaValArgThrArgAsnAsnSerValSerSerArgGluArgHisArgProSer 120  
 Db 415 CCAGTCTGTCGGAACCTCGAATTAACAACAGTGTCTCCAGCGCGGAGAGGACACAGCCCTTCC 474  
 Qy 121 ProArgSerThrArgGlyArgGlnGlyArgGlnGlyArgGlnGlyArgGlnGlyPhe 140  
 Db 475 CCACGTTCCACCCGAGGCGCGGAGGCGGCAACCATGTGGACGAGTCCCGCGGAGTTC 534  
 Qy 141 ProAlaThrArgSerLeuArgArgArgAlaThrAlaSerAlaGlyThrProTrpProSer 160  
 Db 535 CCGGCTACCGAGTCTCCAGCGCGGCGCAACAGCATCGGACGAAACGCAATGCGCCCTCC 594  
 Qy 161 ProProSerSerTyrLeuThrIleAspLeuThrAspAspThrGluAspThrHisGlyThr 180  
 Db 595 CTCTCCAGCTCTTACCTTACCATCGACCTCACAGACGACACAGAGGACACACATGGGACG 654  
 Qy 181 ProGlnSerSerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyMet 200  
 Db 655 CCCCAGAGCAGCAGTACCCCTACGCCCGCTAGCCAGGACGACAGGAGGCGGCGCATG 714  
 Qy 201 GluSerProGlnValGluAlaAspSerGlyAspGlyAspSerSerGluTyrGlnAspGly 220  
 Db 715 GAGTCCCCCGAGTGGAGGACGACACTGGAGATGGAGACATTCAGAGTATCAGGATGG 774  
 Qy 221 LysGluPheGlyIleGlyAspLeuValTrpGlyValIleLysGlyPheSerTrpTrpPro 240  
 Db 775 AAGGAGTTTGGAAATAGGGACCTCTGTGTGGGAAAGATCAAGGCTCTCTCTGTGGGCC 834  
 Qy 241 AlaMetValValSerTrpLysAlaThrSerLysArgGlnAlaMetSerGlyMetArgTrp 260  
 Db 835 GCATGGTGGTGTCTTGGAGGCGCACCTCCCAAGCAGCAGGCTATGTCTGGCATGCGGTGG 894  
 Qy 261 ValGlnTrpPheGlyAspGlyLysPheSerGluValSerAlaAspLysLeuValAlaLeu 280  
 Db 895 GTCCAGTGGTTTGGCGATGGCAAGTCTCCAGGCTCTTCGAGACAACTGGTGGCACTG 954  
 Qy 281 GlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLys 300  
 Db 955 GGGCTCTTCAGCCAGCACCTTTAATTTGGCCACCTTCAATTAAGCTCGTCTCTCTATCGAAA 1014  
 Qy 301 AlaMetTyrHisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSerSer 320  
 Db 1015 GCCATGTACCATGCTCTGGAGAAAGCTAGGCTGGAGCTGGCAAGACCTTCCCCAGCAGC 1074  
 Qy 321 ProGlyAspSerLeuGluAspGlnLeuLysProMetLeuGluTrpAlaHisGlyGlyPhe 340  
 Db 1075 CCTGGAGACTCATTTGAGGACCGAGCTGAAGCCCATGTTGGAGTGGGCGGCGGCTTC 1134  
 Qy 341 LysProThrGlyIleGluGlyLeuLysProAsnAsnThrGlnProValValAsnLysSer 360  
 Db 1135 AAGCCCACTGGGATCGAGGCGCTCAAAACCCAAACACGCAACCAAGTGGTAAATAGTTCG 1194  
 Qy 361 LysValArgArgAlaGlySerArgLysLeuGluSerArgLysTyrGluAsnLysThrArg 380  
 Db 1195 AAGTGGCTGTCGACGAGTAGGAAATTAGAATCAGGAATATCAGGACACAGACTCGA 1254  
 Qy 381 ArgArgThrAlaAspAspSerAlaThrSerAspTyrCysProAlaProLysArgLeuLys 400  
 Db 1255 AGACGCACAGCTGACGACTCAGCCACTCTGACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1314

QY 401 ThrAsnCyTYrAsnAsnGlyLysAspArgGlyAsnGlnSerArgGlnMet 420  
 Db 1315 ACAAAATGCTATTAACAACGCGAAGAGCGAGGAGTGAAGATCAGAGCGGAAACAATG 1374  
 QY 421 AlaSerAspValAlaAsnAsnLysSerSerLeuGlnAspGlyCysLeuSerCysGlyArg 440  
 Db 1375 GCTTCAGATGTTGCCAACAAACAGAGCAGCTGGAAGATGCTGTTGCTGTGGCAGG 1434  
 QY 441 LysAsnProValSerPheHisProLeuPheGlnGlyLysCysGlnThrCysArgAsp 460  
 Db 1435 AAAAACCCTGGTCTTCCACCTCTTTGAGGGGGGCTCTGTCAGACATGCCGGAT 1494  
 QY 461 ArgPheLeuGlnLeuPheTyrMetTyrAspAspAspGlyTyrGlnSerTyrCysThrVal 480  
 Db 1495 CGCTTCCTGAGCTGTTTACATGATGATGAGATGCTATCAGTCTTACACACCTGTG 1554  
 QY 481 CysCysGlnGlyArgGlnLeuLeuLeuCysSerAsnThrSerCysCysArgCysPheCys 500  
 Db 1555 TGCTCCGAGGCGCAGAGAGCTGCTGTTGACAGCAACGAGCTGCTCGGCTTCTGT 1614  
 QY 501 ValGlnCysLeuGlnValLeuValGlyThrGlyThrAlaAlaGlnAlaLysLeuGln 520  
 Db 1615 GTGAGAGTCCCTGAGAGTGTGGTGGCAGAGCAGCGCGCCAGGCTTCAGAGAG 1674  
 QY 521 ProThrSerCysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgLys 540  
 Db 1675 CCCTCGAGCTGCTACATGTGTCTCCGAGCGCTGTGATGCGCTGCGCGCCGAGAG 1734  
 QY 541 AspTrpAsnValArgLeuGlnAlaPhePheThrSerAspThrGlyLeuGlnTyrGlnAla 560  
 Db 1735 GACTGGAACTGTGCGCTGCGAGGCTTCTTACCAAGTACACAGGCGCTTAAATCGAAGCC 1794  
 QY 561 ProLysLeuTyrProAlaIleProAlaAlaArgArgArgProIleArgValLeuSerLeu 580  
 Db 1795 CCCAAGCTGACCTGCTCCATTCGCCAGCGCCGAGGCGCCCATTCGATCTCTGTCATTG 1854  
 QY 581 PheAspGlyIleAlaThrGlyTyrLeuValLeuLysGlnLeuGlyIleLysValGlyLys 600  
 Db 1855 TTTCATGCGATGCGCAGCGCTTACTGCTCAAGAGTTGGGCAATTAAGGTAGGAAAG 1914  
 QY 601 TyrValAlaSerGlnValCysGlnGlnSerIleAlaValGlyThrValLysHisGlnGly 620  
 Db 1915 TACGTGCTTCTGAGAGTGTGAGAGTCCATTGCTGTGGAACCGTGAAGCGAGAGGG 1974  
 QY 621 AsnIleLysTyrValAsnAspValArgAsnIleThrLysLysAsnIleGlnGlnTyrGly 640  
 Db 1975 AATATCAATACGTGACGACGCTGAGAACATCAACAAGAAAATTAATTAAGATGGGGC 2034  
 QY 641 ProPheAspLeuValIleGlyGlySerProCysAsnAspLeuSerAsnValAsnProAla 660  
 Db 2035 CCATTGACTTGATGATTTGGCGGAAGCCATGACAGATCTCTCAATGTGAATCCAGCC 2094  
 QY 661 ArgLysGlyLeuTyrGlnGlyTyrGlyArgLeuPhePheGlnPheTyrHisLeuLeuAsn 680  
 Db 2095 AGGAAAGCCTGTATAGAGGATCAGGCGGCTCTTTCGAATTTTACCACTGCTGAAT 2154  
 QY 681 TyrSerArgProLysGlnGlyAspAspArgProPhePheTyrMetPheGlnAsnValVal 700  
 Db 2155 TACTCAGCCCCCAGAGAGGGATGATACCGGCTCTTCTGATGTTTGAAGATGTTGTA 2214  
 QY 701 AlaMetLysValGlyAspLysArgAspIleSerArgPheLeuGlnCysAsnProValMet 720  
 Db 2215 GCCATGAGGTTGGCGCAAGAGGAGCATCTCAGCTTCCGAGTGTAAATCCAGTATG 2274  
 QY 721 IleAspAlaIleLysValSerAlaAlaHisArgAlaArgTyrPheTyrGlnAsnLeuPro 740  
 Db 2275 ATTGATGACATCAAGATTTTCTGCTGTACAGGGCCGATACTTCTGGGCAACTACCC 2334  
 QY 741 GlyMetAsnArgProValIleAlaSerLysAsnAspLysLeuGlnLeuGlnAspCysLeu 760  
 Db 2335 GGGATGAACAGGCCCGTGTATGATCAAAAGAAATGATAACTCGAGCTGACGACTGCTGTG 2394

QY 761 GluTyrAsnArgIleAlaLysLeuLysValGlnThrIleThrThrLysSerAsnSer 780  
 Db 2395 GAATACATATAGATAGCCAAATTAAAGAAATGACAGACATTAACCAACCAAGTCACTCG 2454  
 QY 781 IleLysGlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysGlnAspValLeu 800  
 Db 2455 ATCAAAACAGGGGAAAAACCACTTTCCTGTTGTATGATGACAAAGAAAGATGTTTG 2514  
 QY 801 TrpCysThrGlnLeuGlnArgGlyPheProValHisTyrThrAspValSerAsn 820  
 Db 2515 TGTGCTAGTACGCTGAAAGATCTTGGCTTCTGTCATGACATCAGACAGCGTCCAAC 2574  
 QY 821 MetGlyArgGlyAlaArgGlnLysLeuLeuGlnArgSerTrpSerValProValIleArg 840  
 Db 2575 ATGGCCCGTGTGCTCCGACAGAACTGCTGGAGAGTCTTGAGACGTGCTGTATCCGA 2634  
 QY 841 HisLeuPheAlaProLeuLysAspTyrPheAlaCysGln 853  
 Db 2635 CACCTCTTCCGCTCTGAAAGACTACTTTCATGTGAA 2673

RESULT 2  
 AF331857 4335 bp mRNA linear PRI 02-JAN-2002  
 LOCUS  
 DEFINITION Homo sapiens DNA cytosine methyltransferase 3 beta (DNMT3B) mRNA,  
 complete cds.  
 ACCESSION AF331857  
 VERSION AF331857.1 GI:18033254  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens.  
 Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE  
 AUTHORS Ni, J., Pradhan, S. and Roberts, R.J.  
 TITLE Cloning, expression and characterization of human DNMT3 genes  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 4335)  
 AUTHORS Ni, J., Pradhan, S. and Roberts, R.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (22-DEC-2000) New England Biolabs, 32 Tozer Road,  
 Beverly, MA 01915, USA

FEATURES  
 source  
 gene  
 CDS  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1..4335  
 /gene="DNMT3B"  
 322..2883  
 /gene="DNMT3B"  
 /codon\_start=1  
 /product="DNA cytosine methyltransferase 3 beta"  
 /protein\_id="AA157040.1"  
 /db\_xref="GI:18033255"  
 /translation="MKGDRILNGEEDAGREDSILVNGACSDQSDSPILIAIRTP  
 EIRGRSSRSIRKREVSILSYTDLTGCGDGDGSDSPMPKLFRETRRESPPA  
 VTRNNNSVSSRRHRPSRSTRGROGRNHVDSPEVERPATRSLRARATSAQTPWS  
 PPSVYLITDITDIEDTHGTPOSSSTPYARLADSDQGMESPOVADLSDGSDSEYO  
 DKEFGIGDLVWGKIKGFSWPMAMVSWATSKRQAMSGMWQVOWGDXGFSYVADK  
 LVALLGFSOHFNATFNKLVSRKAMYHLEKARVRAGKTPESSPDLSLDLKPMLK  
 MAHGAFKPTIGELKPNTPQPVNKSQRKASRKLEBKRYENKTRRTDSDATSDY  
 CPAPRLKTNQYNNKGKRGDEDSRQMASDVANAKSLLEDGLSGRKAPVSHPLF  
 EGGLCOTCDRDFLELYMYDDGYOSYCVCCBGRLEILCSNCCRCFVCELEVLV  
 GTGTABAKLOEPMWSCYMCUPORCGVLRKQMWNLQAFTSDGLGEYAPKLYPA  
 IPAARRPRIRVLSIFDGIATGYLVKELGIVKQKYVASEVESESTAVGTYKHENIKY  
 VNDVRNITRKNIIEWGPPDLVIGSPCNDLSNVNPARGLVEGTGLFEFVYHLNYS  
 RPKEGDDREFPMFENVAMVGDGRDIFELCNPWVITDAIKVSAHARARVFMGNLP  
 GMRNRYVIAKNDLEODCLEYNRJAKLKVOTITTSNLSIKGKQQLFVVMNGKED  
 VLMCTELEIERLIFGPPVHYTIDVSNMGRAGARKLDRSMSPVIRHLFAPDKDYFACE"

BASE COUNT 1045 a 1172 c 1167 g 951 t  
 ORIGIN  
 Alignment Scores: 5.29e-234 Length: 4335  
 Pred. No.:

Score:	4566.00	Matches:	853
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0

US-09-720-086-8 (1-853) x AF331857 (1-4335)

Qy	1	MetLysGlyAspThrArgHisLeuAsnGlyGluAlaGlyArgGluAspSer	20
Db	322	ATGAAGGAGACACACAGGCATCTCAATGGAGAGGAGACCGCGCGGAGGAAGACTCG	381
Qy	21	IleLeuValAsnGlyAlaCysSerAspGlnSerSerAspSerProIleLeuGluAla	40
Db	382	ATCTCGTCAACGGGGCTGCAGCGACAGTCTCCGACTCGCCCCCAATCTCGAGGCT	441
Qy	41	IleArgThrProGluIleArgGlyArgArgSerSerArgLeuSerIleArgGluVal	60
Db	442	ATCCGACACCCCGAGATCAGAGCCGGAAGATCAAGCTCGGACTCTCCNAGAGGAGGTG	501
Qy	61	SerSerLeuLeuSerTyrThrGlnAspLeuThrGlyAspGlyAspGlyValGluAspGlyAsp	80
Db	502	TCCAGTCTGCTAGCTACACACAGGACTTGACAGCGCGATGCGGCGGAGATGGGAT	561
Qy	81	GlySerAspThrProValMetProLysLeuPheArgGluThrArgThrArgSerGluSer	100
Db	562	GGCTCTGACACCCCGAGTCATGCCAAAGCTCTTCCGGGAAACACAGGACTCGTTTCAGAAAGC	621
Qy	101	ProAlaValArgThrArgAsnAsnSerValSerSerArgGluArgHisArgProSer	120
Db	622	CCAGCTGTCCGAACCTCGAATAACAACAGTGTCTCCAGCGGGAGAGCACACAGGCTTCC	681
Qy	121	ProArgSerThrArgGlyArgGlnGlyArgAsnHisValAspGluSerProValGluPhe	140
Db	682	CCACGTTTCCACCGAGCGCGGAGCGCGCAACCATGTGGACGAGTCCCTCCGTGGAGTTC	741
Qy	141	ProAlaThrArgSerLeuArgArgAlaThrAlaSerAlaGlyThrProTrpProSer	160
Db	742	CCGGTCTACAGGTCCTTGAGACGCGCGGCAACAGCATCGGAGGAAACCCATGGCCGTC	801
Qy	161	ProProSerSerTyrLeuThrIleAspLeuThrAspAspThrGluAspThrHisGlyThr	180
Db	802	CCTCCAGCTCTTACCTTACCATCGACCTCACAGCGACACAGAGGACACACATGGGAGC	861
Qy	181	ProGlnSerSerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyMet	200
Db	862	CCCCAGAGCAGCAGTACCCCTACGCCGCTAGCCCCAGGACAGCCAGCGGGGGCATG	921
Qy	201	GluSerProGlnValGluAlaAspSerGlyAspGlyAspSerSerGluTyrGlnAspGly	220
Db	922	GAGTCCCCCGAGGTGGAGGCAGACAGTGGAGATGGAGACAGTTCAGAGTATCAGGATGG	981
Qy	221	LysGluPheGlyIleGlyAspLeuValTrpGlyLysIleLysGlyPheSerTrpTrpPro	240
Db	982	AAGGAGTTTGGAAATAGGGACCTCGTGTGGGAAAGATCAAGGGCTTCTCTGTGGGCC	1041
Qy	241	AlaMetValSerTrpLysAlaThrSerLysArgGlnAlaMetSerGlyMetArgTrp	260
Db	1042	GCCATGGTGTGTCTTGGAAAGGCACCTCCAAAGCGACAGGCTATGTCTGGCATGGCGTGG	1101
Qy	261	ValGlnTrpPheGlyAspGlyLysPheSerGluValSerAlaAspLysLeuValAlaLeu	280
Db	1102	GTCCAGTGTGTGGCGATGGCAAGTTCTCCGAGGTCTCTGCAGACCAACTGGTGGCCTG	1161
Qy	281	GlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLys	300
Db	1162	GGGCTGTTCAGCCAGCATTAAATTTGGCCACCTTCAATAAGCTCGTCTCTATCGAAAA	1221
Qy	301	AlaMetTyrHisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSerSer	320
Db	1222	GCCATGTACCATCTCTGGAGAAAGCTAGGTTGCGAGTGGCAAGACCTTCCCCAGCAGC	1281
Qy	321	ProGlyAspSerLeuGluAspGlnLeuLysProMetLeuGluTrpAlaHisGlyGlyPhe	340

Db	1282	CCTGGAGACTCATTTGGAGGACCACTGAAGCCCATGTTGGAGTGGCCACCGGGGCTTC	1341
Qy	341	LysProThrGlyIleGlyGlyLeuLysProAsnAsnThrGlnProValValAsnLysSer	360
Db	1342	AGCCCACTGGATCGAGGCGCTCAACCCCAACAACACGCAACCACTGTTTAAATAGTCG	1401
Qy	361	LysValArgArgAlaGlySerArgLysLeuGluSerArgLysTyrGluAsnLysThrArg	380
Db	1402	AAGGTGCTGCTGTCAGGCAGTAGGAAATTTAGAAATCAAGAAATACGAGAAACAAGACTCGA	1461
Qy	381	ArgArgThrAlaAspAspSerAlaThrSerAspTyrCysProAlaProLysArgLeuLys	400
Db	1462	AGACGCACAGCTGACGACTCAGCCACCTCTGACTACTGCCCGCCACCCAAAGCGCTCAAG	1521
Qy	401	ThrAsnCysTyrAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGluGlnMet	420
Db	1522	ACAAATTTGCTATACCAACGCGCAAGACCGAGGGATGAAGATCAGAGCGGAGAAACAATG	1581
Qy	421	AlaSerAspValAlaAsnAsnLysSerSerLeuGluAspGlyCysLeuSerCysGlyArg	440
Db	1582	GCTTCAGATGTTGCCAACAACAACAGCAGCCTGGAAAGATGGCTGTTGTCTGTGGCAGG	1641
Qy	441	LysAsnProValSerPheHisProLeuPheGlyGlyLeuCysGlnThrCysArgAsp	460
Db	1642	AAAAACCCCGTGTCTTCCACCTCTCTTTGAGGGGGGGCTCTGTCAAGACATGCCGGGAT	1701
Qy	461	ArgPheLeuGluLeuPheTyrMetTyrAspAspAspGlyTyrGlnSerTyrCysThrVal	480
Db	1702	CGCTTCCTTGAGCTGTTTATCATGTATGATGATGATGCTATCATGCTTACTGCACTGTG	1761
Qy	481	CysCysGluGlyArgGluLeuLeuCysSerAsnThrSerCysCysArgCysPheCys	500
Db	1762	TGTTGGAGGCGCGAGAGCTGCTGCTTTGACAGCAACACAGCTGCTGCCGCTGTTCTGT	1821
Qy	501	ValGluCysLeuGluValLeuValGlyThrGlyThrAlaAlaGluAlaLysLeuGlnGlu	520
Db	1822	GTGAGTGCCTGGAGGTGCTGGTGGGCACAGGCACAGCGCGCGAGGCCAAGCTTCAAGGAG	1881
Qy	521	ProTrpSerCysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgArgLys	540
Db	1882	CCCTGGAGCTGTACATGTGTCTCCCGCAGCGCTGTCTATGCGCTCTGCGGCGCGGAAG	1941
Qy	541	AspTrpAsnValArgLeuGlnAlaPhePheThrSerAspThrGlyLeuGluTyrGluAla	560
Db	1942	GACTGGAACGTGCGCTGCGAGGCTTCTTCCACAGTGACACGGGCTTGAATATGAAGCC	2001
Qy	561	ProLysLeuTyrProAlaIleProAlaAlaArgArgArgProIleArgValLeuSerLeu	580
Db	2002	CCCAAGCTGTACCTGCCATTCCCGCAGCCGAAAGCGCGCCATTCCAGTCTCTGTCATTG	2061
Qy	581	PheAspGlyIleAlaThrGlyTyrLeuValLeuLysGluLeuGlyIleValGlyLys	600
Db	2062	TTTGTGGCATCGGACAGGCTACCTAGTCTCAAGAGTTGGGCATTAAGGTAGGAAAG	2121
Qy	601	TyrValAlaSerGluValCysGluGluSerIleAlaValGlyThrValLysHisGluGly	620
Db	2122	TAGCTCGCTTCTGAAAGTGTGTGAGGAGTCCATTGCTGTGGAAACCGTGAAGCACGAGGG	2181
Qy	621	AsnIleLysTyrValAsnAspValArgAsnIleThrLysLysAsnIleGluGluTrpGly	640
Db	2182	AATATCAATACGTGAACGCGCTGAGGAACATCACAAAGAAAAATATTGAAGATGGGGC	2241
Qy	641	ProPheAspLeuValIleGlyGlySerProCysAsnAspLeuSerAsnValAsnProAla	660
Db	2242	CCATTTGACTTGGTGTATGGCGAAGGCCCATGCAACGATCTCTCAAAATGTGAATCCAGCC	2301
Qy	661	ArgLysGlyLeuTyrGluGlyThrGlyArgLeuPhePheGluPheTyrHisLeuLeuAsn	680
Db	2302	AGAAAGGCGCTGATAGGGGTACAGGCGGGCTCTTCTCGAAATTTTACCACTGCTGTAAT	2361
Qy	681	TyrSerArgProLysGlyAspArgProPhePheTrpMetPheGluAsnValVal	700



Db 2362 TACTCAGCCCCAGAGAGGATGATGACCGCGCTTCTTGATGATGTTGTA 2421  
Qy 701 AlMeltySvalGlyAplyAArgAspIleSerArgPheLeuGluCysAspProValMet 720  
Db 2422 GCCATGAAAGTTGGCCACAAGAGGACATCTCAGCGTTCTCGAGTTGATCCAGATG 2481  
Qy 721 IleAspAlaIleLysValSerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuPro 740  
Db 2482 ATGATGACATGAAAGTTTCTGCTGCTCAGAGGCGCGATCTTGCGGGCAACTACCC 2541  
Qy 741 GlyMetLeuArgProValIleAlaSerLysAsnAspLysLeuGluLeuAspCysLeu 760  
Db 2542 GGGATGAAACAGGCGCTGTGATGATCAAGAAATGATTAACCTGAGCTGAGGACTGCTTG 2601  
Qy 761 GlyTyrAsnArgIleAlaLysLeuLysValGlnThrIleThrLysSerAsnSer 780  
Db 2602 GAATACATAGATAGCCCAAGTTAAAGAAAGTACAGCAATACCAACCAAGTCAACTCG 2661  
Qy 781 IleLysGlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysGluAspValLeu 800  
Db 2662 ATCAAAACAGGGGAAAACCAACTTTCCTGTTGTATGATGCAAGAAAGATGTTTG 2721  
Qy 801 TrpCysThrGluLeuGluArgGlyAArgPheGlyPheProValHisTyrThrAspValSerAsn 820  
Db 2722 TGGTCACTGAGACTCCAAAGATCTTTGGCTTCTGTCGACTACACAGACGTCACAC 2781  
Qy 821 MetGlyArgGlyAlaArgGlnLysLeuGluArgSerTyrPheValProValIleArg 840  
Db 2782 ATGGCGCGTGTGCGCCGCGAGAACTGCTGGGAGAGTCCAGAGGCTGCTGTCATCGA 2841  
Qy 841 HisLeuPheAlaProLeuLysAspTyrPheAlaCysGlu 853  
Db 2842 CACCTCTTGGCCCTTGAAGACTTTCATGATGTA 2880

RESULT 3  
LOCUS AF176228 4267 bp mRNA linear PRI 26-OCT-1999  
DEFINITION Homo sapiens DNA cytosine-5 methyltransferase 3B (DNMT3B) mRNA,  
complete cds.  
ACCESSION AF176228  
VERSION AF176228.1 GI:6118091  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 4267)  
XU, G.-L., Bester, T.H., Bourc'his, D., Hsieh, C.-L., Tommerup, N.,  
Bugge, M., Hulten, M., Qu, X., Russo, J.J., and Viegas-Pequignot, E.  
Chromosome instability and immunodeficiency syndrome caused by  
mutations in a DNA methyltransferase gene  
Nature (1999) In press  
2 (bases 1 to 4267)  
XU, G.-L. and Bester, T.H.  
Direct Submission  
Submitted (06-AUG-1999) Genetics and Development, Columbia  
University, 701 West 168 St., New York, NY 10032, USA  
FEATURES  
source  
1..4267  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="20"  
/map="20q11-q13"  
1..4267  
/gene="DNMT3B"  
/gene="DNMT3B"  
265..2802  
/gene="DNMT3B"  
/note="mutated in human chromosome instability and  
immunodeficiency disease known as ICF syndrome; contains  
alternative 5' exon 1B"  
/product="DNA cytosine-5 methyltransferase 3B"  
/protein\_id="AA04015.1"

/db\_xref="GI:6118092"  
/translation="MESSPEPSSLESKMGDTRHNGERDAGREDSILVNGACSDSS  
DSPPLEAIRTPERIRGRSSRLSRVSSLSLTQDITGDDGDDGSDTPWPKL  
FRRTTRRSPPARTNNNSVSRERHRPSRSTRGGRNHRNDESPVEPATSLR  
RATASAGTPWSPSSVLTIDTDDTDEHGTGOSTTPVALAODSOGMESPOVE  
ADSGDSDSEYOQKXERGIDLVWKIKGSPWPMVPMVSMKATSKROAMSGMPOVE  
GDKRSVNSADKVALGLFSQFNLTATFNKLVSRKAMTHALEKARACKTFFSSPG  
DSLEDDQLEKMLEAHGGEKPTGIEGLKPNNTQENKTRRTADSDATSDYCPAPRLK  
TNCYNGKDRGDEDOSEOMASDVANNKSLBDGSCGRKNPVSFPLFEGGLCOTC  
RDRPLELPMYDDGVQSYCTCCGREGILLCSNTRCCEVECEILVGTGTAAEA  
KLQEPWSCYMCLEPQRCGVLRKKNMNVLAQFTSDTGLYEAPKLYAIPARRRP  
IRVLSLFDGIATGYLVLEKELIKVGRVASEVEESIAGTVKHBEKIKVNDVRNIT  
KKNIEMWGFDFIVIGSPCNDSLVNPARGLVEGTALFEFVHLNLSRPKEGDR  
PFPMPENVVAMRVGDKRDISRPLECNPMVDAIKVSAHRARYFMGNLPGMNPVIA  
SKNDLELDQCLEYNRILAKLKVOTITTSNSIKQGNQLPPVWNGKEVLMCTELE  
RIFGPPVATDVSNMGRGAROKLIGRSWSPVIRHLPAFLDYFACE"

BASE COUNT 1069 a 1098 c 1147 g 953 t  
ORIGIN

Alignment Scores:  
Pred. No.: 7.69e-228 Length: 4267  
Score: 4450.00 Matches: 833  
Percent Similarity: 97.66% Conservative: 0  
Best Local Similarity: 97.66% Mismatches: 0  
Query Match: 97.46% Indels: 20  
DB: 9 Gaps: 1

US-09-720-086-8 (1-853) x AF176228 (1-4267)

Qy 1 MetLysGlyAspThrArgHisLeuAsnGlyGluGluAspAlaGlyValArgGluAspSer 20  
Db 301 ATGAGGAGAGACACAGGACATCTAATGAGAGAGAGACCGCGGGAGGAGACTCG 360  
Qy 21 IleLeuValAsnGlyValaCysSerAspGlnSerSerAspSerProProIleLeuGluAla 40  
Db 361 ATCTCTGTCACAGGGGCTTCGACGACCACTCTCCGACTGCGCCCAANTCTCGAGGCT 420  
Qy 41 IleArgThrProGluIleArgGlyArgArgSerSerArgLeuSerLysArgGluVal 60  
Db 421 ATCCGCAACCCCGAGATCAAGAGCCGAGATCAAGCTCGGACTCTCCAAAGAGGAGGTG 480  
Qy 61 SerSerLeuLeuSerTyrThrGlnAspLeuThrGlyAspGlyAspGlyValAspGlyAsp 80  
Db 481 TCCAGTCTGCTAAGCTACACACAGAGACTTGACAGGCGATGCGGAGCAAGATGGGGAT 540  
Qy 81 GlySerAspThrProValMetProLysLeuPheArgGlyThrArgThrArgSerGlnSer 100  
Db 541 GGCTCTGACACCCCACTCATGCAAGCTCTTCGGGAAACAGAGACTCTTCAGAAAGC 600  
Qy 101 ProLysValArgThrArgAsnAsnAsnSerValSerSerArgGluArgHisArgProSer 120  
Db 601 CCAGCTGTCGAACTCGAANTACACACAGTCTCCAGCCGGAGAGGACAGGCTTCC 660  
Qy 121 ProArgSerThrArgGlyArgGlnGlyArgAsnHisValAspGlnSerProValGluPhe 140  
Db 661 CCAGCTTCCACCCGAGGCGGAGGCGGCAACCATGTGAGACAGTCCCCCGTGGAGTTC 720  
Qy 141 ProLysThrArgSerLeuArgArgAlaThrAlaSerAlaGlyThrProTrpProSer 160  
Db 721 CCGGCTACCAAGTCCCTGAGACGGCGGCAACAGCATCGGACAGAAACGCCATGCGCTCC 780  
Qy 161 ProProSerSerTyrLeuThrIleAspLeuThrAspAspThrGluAspThrHisGlyThr 180  
Db 781 CTCTCCAGCTCTTACTTACCATGACCTCAGACAGACACAGAGACACACATGGGACG 840  
Qy 181 ProGlnSerSerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyMet 200  
Db 841 CCCAGAGCAGAGTAGTACCCCTAGCCCGCTAGCCACAGACACACAGAGGGGGGATG 900  
Qy 201 GluSerProGlnValGluAlaAspSerGlyAspGlyAspSerSerGlyTyrGlnAspGly 220  
Db 901 GAGTCCCGCAGGTGAGGACAGACAGTGAATGAGACAGCTTACAGATACAGATGGG 960

Qy	221	LysGluPheGlyIleGlyAspLeuValTrpGlyLysIleLysGlyPheSerTrpTrpPro	240
Db	961	AAAGGAGTTTGAATAGGGGACCTCGTGTGGGGAAGATCAAGGGGCTTCTCGTGTGGCC	1020
Qy	241	AlaMetValValSerTrpLysAlaThrSerLysArgGlnAlaMetSerGlyMetArgTrp	260
Db	1021	GCATGGTGGTCTTGGAAAGGCCACCTCCAAGCGACAGGCTATGCTGGCATGCGGTGG	1080
Qy	261	ValGlnTrpPheGlyAspGlyLysPheSerGluValSerAlaAspLysLeuValAlaLeu	280
Db	1081	GTCCAGTGGTGTGGCGATGCCAAGTTCTCCGAGGTCTCTGCAGACAAACTGGTGGCACTG	1140
Qy	281	GlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTrpArgLys	300
Db	1141	GGGCTGTTGAGCCAGACACTTAATTTGCCACCTTCAATAAGCTCGTCTCTATCGAAAA	1200
Qy	301	AlaMetTrpHisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSerSer	320
Db	1201	GCATGTACCATGCTCTGGAGAAAGCTAGGTGGAGTGGCAAGACCTTCCCGACGAGC	1260
Qy	321	ProGlyAspSerLeuGluAspGlnLeuLysProMetLeuGluTrpAlaHisGlyGlyPhe	340
Db	1261	CCTGGAGACTCATTTGGAGGACCAAGCTGAAGCCCATGTTGGAGTGGGCCACCGGGGCTTC	1320
Qy	341	LysProThrGlyIleGluGlyLeuLysProAsnAsnThrGlnProValValAsnLysSer	360
Db	1321	AAAGCCACTGGGATCGAGGCGCTCAAAACCAACACGCAACCA-----	1365
Qy	361	LysValArgAlaGlySerArgLysLeuGluSerArgLysTrpGluAsnLysThrArg	380
Db	1366	-----GAGAACAAAGACTCGA	1380
Qy	381	ArgArgThrAlaAspAspSerAlaThrSerAspTrpCysProAlaProLysArgLeuLys	400
Db	1381	AGACGCACAGCTGACACTCAGCCACCTCTGACTACTGCCCGCCACCCAGCGCCCTCAAG	1440
Qy	401	ThrAsnCysTrpAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGluGlnMet	420
Db	1441	ACAAATTTGCTATAAACACCGCAAGACCGAGGGGATGAAGATCAGAGCCGAGAAACAAATG	1500
Qy	421	AlaSerAspValAlaAsnAsnLysSerSerLeuGluAspGlyCysLeuSerCysGlyArg	440
Db	1501	GCTTCAGATGTTGCCAACACACAGACGACGCTGGAAGATGCTGTTGCTTGTGGCAGG	1560
Qy	441	LysAsnProValSerPheHisProLeuPheGluGlyLeuCysGlnThrCysArgAsp	460
Db	1561	AAAAACCCCGTGTCTCCACCTCTCTTTGAGGGGGGGCTCTGTGCAGACATGCCGGGAT	1620
Qy	461	ArgPheLeuGluLeuPheTrpMetTrpAspAspGlyTrpGlnSerTrpCysThrVal	480
Db	1621	CGCTTCTTGAGCTGTTTACATGATGATGACCATGGCTATCAGCTTACTGCACTGTG	1680
Qy	481	CysCysGluGlyArgGluLeuLeuCysSerAsnThrSerCysCysArgCysPheCys	500
Db	1681	TGCTCGAGGGGCGAGAGCTGCTGCTTGCAGCAACACGAGCTGCTGCCGGTGTTCGT	1740
Qy	501	ValGluCysLeuGluValLeuValGlyThrGlyThrAlaAlaGluAlaLysLeuGlnGlu	520
Db	1741	GTGGAGTGCCTGGAGTGCCTGGTGGGCACAGCCACAGCGCGGCGGCAAGCTTCAGGAG	1800
Qy	521	ProTrpSerCysTrpMetCysLeuProGlnArgCysHisGlyValLeuArgArgArgLys	540
Db	1801	CCCTGGAGCTGTACATGTGCTCCCGCAGCGCTGTATGGCGTCTCTCGCGCGCGGAAG	1860
Qy	541	AspTrpAsnValArgLeuGlnAlaPhePheThrSerAspThrGlyLeuGluTrpGluAla	560
Db	1861	GACTGGAACGTGGCGCTGCAGGCTTCTTCCACAGTGACACGGGCTTGAATATGAAGCC	1920
Qy	561	ProLysLeuTrpProAlaIleProAlaAlaArgArgArgProIleArgValLeuSerLeu	580
Db	1921	CCCAAGCTGTACCTGCCATTCGCGAGCGCCGAGGCGGCCCATTCGAGTCTCTGCTATTG	1980
Qy	581	PheAspGlyIleAlaThrGlyTrpLeuValLeuLysGluLeuGlyIleLysValGlyLys	600
Db	1981	TTTGATGGCATCGCAGAGGCTACCTAGTCTCAAGAGTTGGGCATAAAGGTAGGAAAG	2040
Qy	601	TrpValAlaSerGluValCysGluGluSerIleAlaValGlyThrValLysHisGluGly	620
Db	2041	TACGTGCTTCTGAAGTGTGTGAGGAGTCCATGCTGTGTGGAACCGTGAAGCACGAGGGG	2100
Qy	621	AsnIleLysTrpValAsnAspValArgAsnIleThrLysLysAsnIleGluGluTrpGly	640
Db	2101	AATATCAATACGTGAACGACGCTGAGGAACATCAAAAGAAAAATATTGAAGAATGGGGC	2160
Qy	641	ProPheAspLeuValIleGlyLysProCysAsnAspLeuSerAsnValAsnProAla	660
Db	2161	CCATTTGACTTGGTATTGGCGAAGCCCATGCAACGATCTCTCAATGTGAATCCAGCC	2220
Qy	661	ArgLysGlyLeuTrpGluGlyThrGlyValArgLeuPheGluPheGlyTrpHisLeuAsn	680
Db	2221	AGGAAGGCTGTATGAGGGTACAGGCGGCTCTTCTCGAATTTTACCACCTGCTGAAT	2280
Qy	681	TrpSerArgProLysGluGlyAspArgProPhePheTrpMetPheGluAsnValVal	700
Db	2281	TACTCACGCCCAAGAGGAGGTGATCACCGGCGTCTTCTCTGGATGTTTGAGATGTTGTA	2340
Qy	701	AlaMetLysValGlyAspLysArgAspIleSerArgPheLeuGluCysAsnProValMet	720
Db	2341	GCCATGAGGTTGGCGACAGAGGACATCTCACGGTCTCTGGAGTGAATTCAGTGATG	2400
Qy	721	IleAspAlaIleLysValSerAlaAlaHisArgAlaArgTrpPheTrpGlyAsnLeuPro	740
Db	2401	ATTGATGCCATCAAAAGTTTCTGCTGCTCACAGGCGCCGATACTTCTGGGCAACCTACCC	2460
Qy	741	GlyMetAsnArgProValIleAlaSerLysAsnAspLysLeuGluLeuGlnAspCysLeu	760
Db	2461	GGATGACAGGCGCTGATAGCATCAAGAAATGATAAATCGAGCTGCAGACTGCTTG	2520
Qy	761	GluTrpAsnArgIleAlaLysLeuLysValGlnThrIleThrLysSerAsnSer	780
Db	2521	GAATACAAATAGGATAGCCAAAGTTAAAGAAAGTACAGACAATAACCAAGTCGAACTCG	2580
Qy	781	IleLysGlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysGluAspValLeu	800
Db	2581	ATCAACAGCGGAAAAACCAACTTTTCCCTGTTGTCATGAATGGCAAGAGATGTTTG	2640
Qy	801	TrpCysThrGluLeuGluArgIlePheGlyPheProValHisTrpThrAspValSerAsn	820
Db	2641	TGTTGACTGAGCTCGAAGAGATCTTGGCTTTCTGTCGACTACAGACGTTGCTCAAC	2700
Qy	821	MetGlyArgGlyAlaArgGlnLysLeuLeuGlyArgSerTrpSerValProValIleArg	840
Db	2701	ATGGGCGTGTGTCGCCCGCAGAGCTGCTGGGAAGTCTCTGGAGCGTCTGTCATCCGA	2760
Qy	841	HisLeuPheAlaProLeuLysAspTrpPheAlaCysGlu	853
Db	2761	CACCTCTTCGCCCTCTGAAGGACTACTTTGTCATGTGAA	2799
RESULT	4		
AF156487			
LOCUS			
DEFINITION		3897 bp mRNA linear PRI 05-SEP-1999	
		Homo sapiens DNA cytosine-5 methyltransferase 3 beta 3 (DNMT3B)	
ACCESSION		mRNA, complete cds.	
VERSION		AF156487	
KEYWORDS		AF156487.1 GI:5823165	
SOURCE			
ORGANISM			
		Homo sapiens	
		Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
		Mammalia; Eutheria; Primates; Carnivora; Hominidae; Homo.	
REFERENCE		1 (bases 1 to 3897)	
AUTHORS		Xie S., Wang Z., Okano M., Nogami M., Li Y., He W.W., Okumura K.	
		and Li E.	
TITLE		Cloning, expression and chromosome locations of the human DNMT3	
JOURNAL		Gene 236 (1), 87-95 (1999)	

MEDLINE	99365304
PUBMED	10433969
REFERENCE	2 (bases 1 to 3897)
AUTHORS	Xie, S. and Li, E.
TITLE	Direct Submission
JOURNAL	Submitted (03-JUN-1999) CVRC, Mass. Gen. Hospital, 149 13th Street Charlestown, MA 02129, USA
FEATURES	Location/Qualifiers

BASE COUNT	951 a	1039 c	1033 g	874 e
ORIGIN				

	Alignment Scores:	
Pred. No.:	2.17e-208	Length: 3897
Score:	4083.50	Matches: 770
Percent Similarity:	90.27%	Conservative: 0
Best Local Similarity:	90.27%	Mismatches: 0
Query Match:	89.43%	Indels: 83
DB		Gaps: 2
US-09-720-086-8 (1-853)	x AF156487 (1-3897)	

QY	1	MeLlyesGIyAspThrArgHISleuAlaNGlyGluGluAspAlaGlyGlyArgGluAspSer	20
Db	115	ATGAGGGAGACACCAAGCATCTCAATGAGAGAGAGACGCCGGGCGGGAGGAGACGCTG	174
QY	21	ILeLeuValaAmGlyAlaCysSerAspGlnSerSerAspSerProPoiLeuGluAla	40
Db	175	ATCCTCGTCAACGGGGCCCTGCAGGAGCAAGTCTCCGACTCGCCCCCATCTCGAGGCT	234
QY	41	ILeArgThrProGluIleArgGlyArgArgSerSerArgLeuSerLysArgGluVal	60
Db	235	ATCCGACACCCGGAGATCAGAGGCCGAAGATCAAGCTCGGACTCTCCAAAGGAGGGTG	294
QY	61	SeSerSerLeuLeuSerTyrThrGlnAspLeuThrGlyAspGlyAspGlyGluAsp	80
Db	295	TTCAGTGTGCTAAAGCTACACACAGGACTTGACAGCGGATGGCGACGGGGAAGATGGGAT	354
QY	81	GlySerAspThrProValMetProLysLeuPheArgGluThrArgThrArgSerGluSer	100
Db	355	GGCTCTACACCCCAAGTCAATGCCAAGGCTCTCCGGGAAACCAAGGACTCGTTCACAAAC	414
QY	101	ProAlaValArgThrArgAsnAsnAsnSerValSerSerArgGluArgHisArgProSer	120
Db	415	CGAGCTGTCGACACTGGAATTAACACAGTGTCTCCAGCCGGGAGAGGCACAGGCTTTC	474

Oy	121	ProARSerThrArgGlyValArgInglValArgAsnHisValAAspGluSerProValGluPhe	140
Db	475	CCACGTTCCACCGAGGCGGAGGCGGCAACCATGTGCAGACTCCCCGGAGACTTC	534
Oy	141	ProAlaThrArgSerLeuArgArgArgAlaThrAlaSerAlaGlyThrProTyrProSer	160
Db	535	CCGGCTACCAAGTCCCTTAGAGAGCGCGGCAACAGCATTCGAGGAACGCGATGCGCTCC	594
Oy	161	ProProSerSerTyrLeuThrThrIleAspLeuThrAspAspThrGluAspThrHisGlyThr	180
Db	595	CTCCACGCTTACCTTACCTTACCATTCGACTTCACAGCAGCACAGAGAGACACACTGGGACG	654
Oy	181	ProGluSerSerSerThrProTyrAlaArgLeuAlaGluAspSerGlnInglValMet	200
Db	655	CCCCAGAGCAGACGATACCCCTTACCGCCGCTAGCCAGCAGCAGCAGAGGCGGGCATG	714
Oy	201	GluSerProGlnValGluAlaAspSerGlyAspGlyAspSerSerGluTyrGlnAspGly	220
Db	715	GAATCCCCGAGGAGGAGGAGCAGACACTGGAGATGGAGCAGCTTCAGATCAGAGATCGAGAGGG	774
Oy	221	lysGluPheGlyIleGlyAspLeuValTyrGlyValIleGlyPheSerTyrPro	240
Db	775	AAGAGATTGGAAATAGGGAGCCTCGTGGGGAAGATCAAGGGGCTTCTCCGTGGTGCC	834
Oy	241	AlaMetValValSerTyrPheValaThrSerLeuArgGlnAlaMetSerGlyMetArgTyr	260
Db	835	GCCATGGTGGTGTTCGTGGAGGCCACTCCAAAGCAGAGGCTATGTCTGGCATGGCGGTGG	894
Oy	261	ValGluTyrPheGlyAspGlyLysPheSerGluValSerAlaAspLysLeuValAlaLeu	280
Db	895	GTCAGTGTTTGGCGAGGCGCAAGTCTCCGAGGTCTCTGACAGACAACTGGTGCACTG	954
Oy	281	GlyLeuPheSerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLys	300
Db	955	GGGCTGTTCAAGCCAGCACCTTTAATTTGGCCACTTCATATAGCTGCTCTTCCTTACGAAA	1014
Oy	301	AlaMetTyrHisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSerSer	320
Db	1015	GCCATGACCATGCTCTCGAGAAAGTATGGGTGGAGGTGGCAACAACCTTCCCCAGCAGC	1074
Oy	321	ProGluAspSerLeuGluAspGlnLeuLysProMetLeuGluTyrAlaHisGlyGlyPhe	340
Db	1075	CTCGAGACCTCATTTGGAGAGCAGCAGCGAAGCCCATGTTGGAGTGGGCCACCGGGGGCTTC	1134
Oy	341	LysProThrGlyIleGluGlyLeuLysProAsnAsnThrGlnProValValAsnLysSer	360
Db	1135	AAGCCCATCGGGATCGAGGGGCTCAAAACCAACAACGCAACCA-----	1195
Oy	361	LysValArgArgAlaGlySerArgLysLeuGluSerArgLysTyrGluAsnLysThrArg	380
Db	1180	-----GAGAAACAAGACTCGA	1199
Oy	381	ArgArgThrAlaAspAspSerAlaThrSerAspTyrCysProAlaProLysArgLeuLys	400
Db	1195	AGAGCGCACACTGACGACTCAGCCACTCTGACTACGCCCCGCAACCCACTCAAG	1254
Oy	401	ThrAsnCysTyrAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGluGlnMet	420
Db	1255	ACAAATTGCTATAACAAACGGCAAGACCGAGGGATGAAAGATCAAGCCGAGAACAATAG	1314
Oy	421	AlaSerAspValAlaAsnAsnLysSerSerLeuGluAspGlyCysLeuSerCysGlyArg	440
Db	1315	GCTTCAGATGTTGGCAACAACAAGACGAGCCTGGAAATGCTGTTTGTCTGTGGCAGG	1374
Oy	441	LysAsnProValSerPheHisPheLeuPheGluIleGlyLeuCysGlnThrCysArgAsp	460
Db	1375	AAAAACCCCGTGTCTCTTCCACCCCTCTTTGAGAGGGGGGCTCTGTACAGCATGCGGGAT	1434
Oy	461	ArgPheLeuGluLeuPheTyrMetTyrAspAspAspGlyTyrGlnSerTyrCysThrVal	480
Db	1435	CGCTTCCCTTGAAGCTGTTTACATGATATGATGACGATGGCTATCACTTCACTGCACTGTG	1494
Oy	481	CysCysGluGluValArgGluLeuLeuLeuCysSerAsnThrSerCysCysArgCysPheCys	500

Db 2386 CACCTCTTCGCCCTCTGAAGGACTACTTTCGATGTGAA 2424

RESULT 5  
AF068626  
LOCUS  
DEFINITION  
AF068626 4195 bp mRNA linear ROD 06-DEC-1999  
Mus musculus DNA cytosine-5 methyltransferase 3B1 (Dnmt3b) mRNA,  
alternatively spliced, complete cds.

ACCESSION  
AF068626  
KEYWORDS  
AF068626.2 GI:6449469  
SOURCE  
Mus musculus.  
ORGANISM  
Mus musculus  
Taxonomy: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 4195)  
Okano,M., Xie,S. and Li,E.  
Cloning and characterization of a family of novel mammalian DNA  
(cytosine-5) methyltransferases  
JOURNAL Nat. Genet. 19 (3), 219-220 (1998)  
MEDLINE 98324766  
PUBMED 9662389  
REFERENCE  
2 (bases 1 to 4195)  
Xie,S., Okano,M. and Li,E.  
Direct Submision  
TITLE Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,  
Charlestown, MA 02129, USA  
JOURNAL  
REFERENCE  
3 (bases 1 to 4195)  
Okano,M., Chijiwa,T., Sasaki,H. and Li,E.  
Direct Submision  
TITLE Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,  
Charlestown, MA 02129, USA  
JOURNAL  
REMARK  
COMMENT Sequence update by submitter  
FEATURES  
On Nov 18, 1999 this sequence version replaced gi:3327979.  
source  
1..4195  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/map="84.0 CM"  
1..4195  
/gene="Dnmt3b"  
misc\_feature  
1..4195  
/gene="Dnmt3b"  
note="similar to EST sequences deposited in GenBank  
Accession Numbers AA16694, AA119979, AA177277, AA210568  
AA407106, and AA575617"  
269..2848  
/gene="Dnmt3b"  
/function="de novo DNA methylation"  
note="alternatively spliced product; contains Cys-rich  
region; C-terminal region is similar to corresponding  
region of DNA cytosine-5 methyltransferase 3A"  
/codon\_start=1  
/product="DNA cytosine-5 methyltransferase 3B1"  
/protein\_id="AAC40178.2"  
/db\_xref="GI:6449470"  
translation="MKGDSRHLNEEGAGSYEECIIVNGNFSQSSDTKDAPSPPVLE  
AICPEVCTPTRGRSSRLKEVSSLLNYTQDMTGDGRDDDEVDDNGSDILMPK  
LTRTKDTRTRESAPVTRHNGSTSLERQASPRITRGRQGRHHVQEYVPVEFPATR  
SRRRRASASTPMSPASVDPMEEVTPKSVSPVDSIQDQDQDMTQVDAASDR  
GSTEYQDKFQIGDILGVNGKIKRQSWPAMVYSWKATSKRQAMPGRMWQVFGDGK  
SEIADKLVALGLFQHNFLATFNKLVSRYKAMVHTLEKARVAKGTSSSPGSESLD  
QKPMLEWAHGFKPTGIEGLKPKQPVNKSVRSDSNLPPRRKNSRRRTTN  
DSASESPPEKRLTSGQKGRGDEEERPMASVETNNKGNLEDCSCGKNPV  
PHPIFEGGLCSOCDRFLFYWBDEQVQSYCTVCCGREGELLCSNTSCRCFCVBC  
LEVLVAGTAGADAKQEPWCNCLPQRCHGVLRKRDNNRLQDFFITDIDLEFEP  
PKLPAIPAKRRPRLVLSLFDGIATGYLVKLGLIKVEKTIASEVCAESIAVGTVFK  
EQGIKYNDVNRKRTIKNLEEWGPFDLVIGSPCNDLSVNPAPKGLYEGTGRLLFFEY  
HLLNYPTRPEGDNRPFFWFWNVAMVKNDKDISFLACNPVMIDAIKVSAARHARY  
FWGNLPGNRPVMAKNDKLEQDCLEFSTAKLKKVQITTKSGNSIRQGNQLPFPV  
MNGKDDVLWCTCEITFIFFGPAHYTDVSNMGRGARQKLLGRSWSVPVINHLFAPLKDPY  
ACE"

BASE COUNT 1061 a 1059 c 1125 g 950 t

## ORIGIN

## Alignment Scores:

Pred. No.:	4,33e-191	Length:	4195
Score:	3759.00	Matches:	717
Percent Similarity:	88.63%	Conservative:	55
Best Local Similarity:	82.32%	Mismatches:	69
Query Match:	82.33%	Indels:	30
		Gaps:	10

US-09-720-086-8 (1-853) x AF068626 (1-4195)

```

QY 1 MetLysGlyAspThrArgHisLeuAsnGlyGluGlnAspAlaGlyGlyArgGluAspSer 20
Db 269 ATGAAGGAGACAGCAGACATCTGAATGAAGAAGAGGCTGCCAGCGGATATGAGAGTGC 328
QY 21 IleuValAsnGlyAlaCysSerAspGlnSerSerAsp-----SerPro 35
Db 329 ATTATCGTTAATGGAACTTCACTGACCACTCTCAGACAGAGAGATGCTCCCTCACCC 388
QY 36 ProIleuGluAlaIle-----ArgThrProGluIleArgGlyArgArg 50
Db 389 CCACTCTTGGAGGCATCTGCACAGAGCCAGCTGCACACAGAGACCGAGGCGCAGG 448
QY 51 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuSerTyrThrGlnAspLeu 70
Db 449 TCAAGCTCCCGCTGTCTTAAGAGGAGGTCTCCAGCCTTCTGAATTACCGCAGGACATG 508
QY 71 ThrGlyAspGlyAspGly-----GluAspGlyAspGlySerAspThrProVal 86
Db 509 ACAGAGAGATGGAGACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 565
QY 87 MetProLysLeuPheArgGlu-----ThrArgThrArgSerGlnSerProAlaVal 103
Db 566 ATGCCAAAGCTCACCCCTGAGACCAAGAGACACAGAGCGCGCTCGAAAGCCCGGCTGTC 625
QY 104 ArgThrArgAsnAsnAsnSerValSerSerArgGluArgHisArgProSerProArgSer 123
Db 626 CGAACCCGACATAGCAATGGAGACCTCCAGCTTGGAGAGGCAAGAGCCCTCCCCAGATC 685
QY 124 ThrArgGlyArgGlnGlyArgAsnHisValAspGluSerProValGluPheProAlaThr 143
Db 686 ACCCGAGGTGGGAGGCGCCGACCATGTGCAAGATACCTGTGAGATTCCGGCTACC 745
QY 144 ArgSerLeuArgArgArgAlaThrAlaSerAlaGlyThrProTyrProSerProSer 163
Db 746 AGGTCTGGAGAGCTGAGGATCGTCTTCAAGACAGCGCATGTGATCCCTGCCAGC 805
QY 164 SerTyrLeuThrIleAspLeuThrAspAspThrGluAspThrHisGlyThrProGlnSer 183
Db 806 -----GTCCACTTCATGAGAGAGTG-----ACACTTAAGAGC 838
QY 184 SerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyMetGlnSerPro 203
Db 839 GTCACTACCCCATCAGTTGACTTGAAGCCAGATGAGATCAGAGAGGTATGATACACCA 898
QY 204 GlnValGluAlaAspSerGlyAspGlyAspSerSerGlnTyrGlnAspGlyLysGluPhe 223
Db 899 CAGGTGATGACAGAGACAGATGAGATGAGACAGACAGATATCAGATATGATTAAGAGTTT 958
QY 224 GlyIleGlyAspLeuValTyrGlyLysIleLysGlyPheSerTyrTyrProAlaMetVal 243
Db 959 GGAATAGGTGACCTCGTGGGGAAGATCAAGGCTTCTCCTGTGGCGCTGCATGCTG 1018
QY 244 ValSerTyrLysAlaThrSerLysArgGlnAlaMetSerGlyMetArgTyrValGlnTyr 263
Db 1019 GTGTCTGGAAAGCCACCTCCAGCGACAGGCGCATGCCCGGAATGGCTGGATGACAGTGG 1078
QY 264 PheGlyAspGlyLysPheSerGluValSerAlaAspLysLeuValAlaLeuGlyLeuPhe 283
Db 1079 TTTGGTGTATGGCAAGTTTCTGAGATCTCTGTGACAAACTGTGGCTCGGGGCTGTC 1138
QY 284 SerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLysAlaMetTyr 303

```

```

Db 1139 AGCCAGCACTTAAATGTGCTACTTCAATTAAGCTGTTCTTATAGAAAGCCCATGTC 1198
QY 304 HisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSerSerProGlyAsp 323
Db 1199 CACACTCTGGAGAAAGCCAGGTTTCGAGCTGGCAGACCTTCTCCAGAGATCTTGGAGAG 1258
QY 324 SerLeuGluAspGlnLeuLysProMetLeuGlnTyrAlaHisGlyGlyPheLysProThr 343
Db 1259 TCACTGGAGAGACCACTGAAAGCCCATGCTGAGTGGGCCCCAGGCTGCTTCAAGCTTACT 1318
QY 344 GlyIleGlyLysLeuLysProAsnAsnThrGlnProValAlaAsnLysSerLysValArg 363
Db 1319 GGGATTCGAGGCTTCAAAACCCCAAGAGACCAACAGTGTGTTAATTAAGTCAAGTGCCT 1378
QY 364 ArgAlaGlySerArgLysLeuGluSerArgLysTyrGluAsnLysThrArgArgArgThr 383
Db 1379 CGTTCCAGACAGTACGAACCTTAGAACCCAGAGACCGGAGAACAAAGTCGAGACCGACA 1438
QY 384 AlaAspAspSerAlaThrSerAspTyrCysProAlaProLysArgLeuLysThrAsnCys 403
Db 1439 ACCAATGACTCTGCTGCTTGTAG--TCCCCCCCAAGCCCTCAAGACAAATAGC 1495
QY 404 TyrAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGlnMetAlaSerAsp 423
Db 1496 TAT--GGCGGGAAGACCGAGGAGAGATGAGAGAGACCGAGAACGATGCTTCTGAA 1552
QY 424 ValAlaAsnAsnLysSerSerLeuGluAspGlyCysLeuSerCysGlyArgLysAsnPro 443
Db 1553 GTCAACCAACAAGGCGCATCTGGAAGACCCGCTGTTGCTGTGGAAGAAAGAACCT 1612
QY 444 ValSerPheHisProLeuPheGluGlyLysLeuCysGlnThrCysArgAspArgPheLeu 463
Db 1613 GTGCTCTTCCACCCCTCTTGAAGGTGGGCTGTGCAGAGTTGCGGGATGCTTCTTA 1672
QY 464 GluLeuPheTyrMetTyrAspAspAspGlyTyrGlnSerTyrCysThrValCysCysGlu 483
Db 1673 GAGCTCTTCTACATGATGATGATGAGAGCGCTATCAGTCTACTGACCGTGTGCTGAG 1732
QY 484 GlyArgGluLeuLeuLeuCysSerAsnThrSerCysArgCysPheCysValGluCys 503
Db 1733 GGCGGTGAACCTGCTGTGACAGTACAGAACAGCTGCTGACAGATGCTTCTGTGGAGTGT 1792
QY 504 LeuGluValLeuValGlyThrGlyThrAlaAlaGluAlaLysLeuGlnLysProTyrSer 523
Db 1793 CTGAGAGTGTGTTGGGCGGAGGACAGCTGAGATGCCAAGCTGACAGAAACCTTGAGC 1852
QY 524 CysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgArgLysAspTyrAsn 543
Db 1853 TGCTATATGTGCTCCTCCCTCAGCGCTGCCATGGGGTCTCCGACCGAGAAAGATTGGAAC 1912
QY 544 ValArgLeuGlnAlaPhePheThrSerAspThrGlyLeu---GluTyrGluAlaProLys 562
Db 1913 ATGGCGCTGACAGACTTCTTCACTACTGATCTGACCTGGAAAGATTGAGCACCAAG 1972
QY 563 LeuTyrProAlaIleProAlaAlaArgArgArgProIleArgValLeuSerLeuPheAsp 582
Db 1973 TTGTATCCAGCAATCTCTGACGACCAAAAGAGGCGCATATGAGAGTCTGCTGTGTTGAT 2032
QY 583 GlyIleAlaThrGlyTyrLeuValLeuLysGluLeuGlyIleLysValGlyLysTyrVal 602
Db 2033 GGAATTGGCAACGGGGTACTTGGTCTCAAGAGCTTGGTATTTAAAGTGGAAAGTACATT 2092
QY 603 AlaSerGluValCysGluGluSerIleAlaValGlyThrValLysHisGluGlyAsnIle 622
Db 2093 GCCTCCGAAGTGTGTGACAGATTCATCGCTGTGGAACTGTTAAGCATGAGAGCCAGATC 2152
QY 623 LysTyrValAsnAspValArgAsnIleThrLysLysAsnIleGluGluTyrGlyProPhe 642
Db 2153 AAATATGTCAATGACGTCCCGAAATCAACAAAGAAATATTTGAAGAGGGGCGCGCTTC 2212
QY 643 AspLeuValIleGlyLysSerProCysAsnAspLeuSerAsnValAsnProAlaArgLys 662

```

Db 2213 GACTTGGTGATGGTGAAGCCCAATGATCTCTCTAACGTCATCTCCCGCGCAAA 2272  
Qy 663 GlyLeuTyrGluGlyThrGlyArgLeuPhePheGluPheTyrHisLeuLeuAsnTyrSer 682  
Db 2273 GOTTATATAGGGGACAGAGGCTCTCTTCGAGTTTACCACTTGTCTGAATATATACC 2332  
Qy 683 ArgProLysGluGlyAspArgProPhePheTyrMetPheGluAsnValValalaMet 702  
Db 2333 CCCCCAAGAGGGGCGAACCGTCCATCTCTTCGAGTATTCGAGAAATGTTGTGGCCATG 2392  
Qy 703 LysValGlyAspLysArgAspLysSerArgPheLeuGluCysAsnProValMetIleAsp 722  
Db 2393 AAAGTGAATGACAAGAAAGACATCTCAAGATTCTCGCATGTAAACCCAGTCATGATCAT 2452  
Qy 723 AlalLeuValSerAlalaHisArgAlalaArgTyrPheTyrGlyAsnLeuProGlyMet 742  
Db 2453 GCCATCAAGGTGCTCTGCTCTACAGGGCCCGGTACTTCTGGGGTAACCTACCCCGGAATG 2512  
Qy 743 AsnArgProValIleAlaSerLysAsnAspLysLeuGluLeuGlnAspCysLeuGluTyr 762  
Db 2513 AACAGGCCGCTGATGCTTCAAGAATGATAAGCTCGAGCTGCAGGACTGCTGGAGTTC 2572  
Qy 763 AsnArgIleAlaLysLeuLysValGlnThrIleThrLysSerAsnSerIleLys 782  
Db 2573 AGTAGACAGCAAGTTAAAGAAGTGCAGACATATACCACTGCACTCCATCAGA 2632  
Qy 783 GlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysGluAspValLeuTyrCys 802  
Db 2633 CAGGCAAAACACAGCTTTTCCCTGTAGTCATGAATGCGAAGGACGAGTTCCTGGTGC 2692  
Qy 803 ThrGluLeuGluArgIlePheGlyPheProValHisTyrThrAspValSerAsnMetGly 822  
Db 2693 ACTGAGCTCGAAGGATCTTCGCTCTCCCTGCTCACTACACGGAGCTGTCACATGGGC 2752  
Qy 823 ArgGlyAlaArgGlnLysLeuGlyArgSerTrpSerValProValIleArgHisLeu 842  
Db 2753 CCGCGGCGCCGTCAGAGCTGCTGGGCGAGTCTCTGGAGTGTACCGGTATCAGACACCTG 2812  
Qy 843 PheAlaProLeuLysAspTyrPheAlaCysGlu 853  
Db 2813 TTTGCCCTTGAAGGACTACTTTGCTGTGAA 2845

RESULT 6  
AF151973 4223 bp mRNA linear ROD 08-JUN-2000  
LOCUS Mus musculus DNA cytosine-specific methyltransferase isoform 5  
DEFINITION AF151973 (Dnmt3b) mRNA, complete cds.  
ACCESSION AF151973  
VERSION AF151973.1 GI:8347127  
KEYWORDS Mus musculus.  
SOURCE Mus musculus.  
ORGANISM Mus musculus.  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
AUTHORS Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and Shen, Y.  
TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing isoforms in mouse embryonic tissue  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 4223)  
AUTHORS Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and Shen, Y.  
TITLE Direct Submission  
JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular Biology, Institution of Basic Medical Sciences, Chinese Academy of Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005, P.R.China  
FEATURES Location/Qualifiers  
source 1..4223  
/organism="Mus musculus"  
/strain="KM"  
/db\_xref="taxon:10090"

/dev\_stage="8-9 day old embryo"  
1..4223  
/gene="Dnmt3b"  
297..2876  
/gene="Dnmt3b"  
/function="methylates cytosine in DNA"  
/note="alternatively spliced"  
/codon\_start=1  
/product="DNA cytosine-specific methyltransferase isoform 5"  
/protein\_id="AAF74519.1"  
/db\_xref="GI:8347128"  
/translation="MKGDSRLNREEGASGYEECIIVNGNFSDDSDTKDAPSPVLE  
AICTPVCPTETRESRSRLSKREYSLNLTQDMTGDGDRDDEVDDGSDILMPK  
LTREKDTETRESPAVTRTHSNGTSLERQASPRITRGQGRHHVQEPVEFPATR  
SRRASASASPWSVSDVDFEETVPKSVTPSVLSDQDQEGMTTQVDAESI  
GSTYIQDDKEFGIGLVNKGKIPWPAWVSWKATSKRAKARVAGTSSSGESL  
SEISADKLVALGLFSQHNLTATNKLIVYRKAMTHLLEKARVAGTSSSGESL  
QUKPMLEWAGGFKPTGIEGLKPNKQPVVNKSVRSDSNLEPRRNRKRRRTTN  
DSAAESPPPKLKTNSYGGKDEDEERMASEVTNNKGLLEDCLSCCKNPVS  
FLFEGGLQCQRDLFLFYMYDEGYQSYCTVCCCEGRELLELLCSNCCFCVCE  
LELVAGATAEADKLOEPWSCVCLPQCHGVLRKRRKDMNRLQDFTTDPLEBEP  
KLYPAI PAARRPIRVLSLFDGIATGLVLKELGIVKVEYIASEVCAESAVGTVKH  
EQIKYVNDVRKTKNIIEWGPDLVIGGSPCNLSNVNPARGLYEGTGLRLEFEPY  
HLINYPKAGDNRPFFMFEVNVKMDKKDISRLACNPMDALIKVSAHARAY  
PWGNLPGMNRPMASIKNDKLELQDCLFEFRTAKLKVOTITTKNSIRQGNQLFPVV  
MNGKDDVLWCTELERIFGPHYITDVSNMGRGAKQKLGWSVPIRHLFAPLKDIFY  
ACE"

BASE COUNT 1051 a 1084 c 1147 g 941 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.68e-190 Length: 4223  
Score: 3748.00 Matches: 716  
Percent Similarity: 88.52% Conservative: 55  
Best Local Similarity: 82.20% Mismatches: 70  
Query Match: 82.08% Indels: 30  
DB: 10 Gaps: 10

US-09-720-086-8 (1-853) x AF151973 (1-4223)

Qy 1 MetLysGlyAspThrArgHisLeuAsnGlyGluAspAlaGlyGlyArgGluAspSer 20  
Db 297 ATGAAGGGAGACAGACACATCTGAATGAAGAAGAGGGTGCAGCGGGTATGAGGAGTGC 356  
Qy 21 IleLeuValAsnGlyAlaCysSerAspGlnSerSerAsp-----SerPro 35  
Db 357 ATTATCGTTAATGGAACTTCAGTGACCACTTCAGACACAGAGGATGCTCCCTCACCC 416  
Qy 36 ProIleLeuGluAlaIle-----ArgThrProGluIleArgGlyArgArg 50  
Db 417 CCAGTCTTGAGGCAATCTGCACAGAGCCAGTCTGCACACAGACAGAGCCGCGCAGG 476  
Qy 51 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuSerTyrThrGlnAspLeu 70  
Db 477 TCNAGTCCCGGTGTCTAAGAGGGAGGTCTCCAGCCTTCTGAATTACAGCAGGACATG 536  
Qy 71 ThrGlyAspGlyAspGly-----GluAspGlyAspGlySerAspThrProVal 86  
Db 537 ACAGGAGATGGAGACAGAGATGATGAAGTAGATGATGGAAATGGCTCTGATATT---CTA 593  
Qy 87 MetProLysLeuPheArgGlu-----ThrArgThrArgSerGluSerProAlaVal 103  
Db 594 ATGCCAAAGCTCACCCGTGAGACCAAGGACACAGGACGCGCTCTGAAAGCCCGCGTGC 653  
Qy 104 ArgThrArgAsnAsnAsnSerValSerArgGluArgHisArgProSerProArgSer 123  
Db 654 CGAACCCGATAGCAATGGGACCTCCAGCTTGAGAGGCAAGAGCCTCCCCCAAGATC 713  
Qy 124 ThrArgGlyArgGlnGlyArgAsnHisValAspGluSerProValGluPheProAlaThr 143  
Db 714 ACCCGAGGTGGCAGGGCCGCCACCATGTGCAGGAGTACCTGTGTGGAGTTTCCGGTACC 773



QY 144 ArgSerLeuArgArgAlaThrAlaSerAlaGlyThrProTrpProSerProSer 163  
DB 774 AGGCTCGAGAGCGTCGAGCATCGCTTCAGCAAGCCGCGATGTCATCCCTGCCAGC 833  
QY 164 SerTyrLeuThrIleAspLeuThrAspSerThrGlyAspThrHisGlyThrProGlnSer 183  
DB 834 -----GTCCACTTCATGAGAAAGTC-----ACACCTTAAGAGC 866  
QY 184 SerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyMetGluSerPro 203  
DB 867 GTGAGTACCCCATCAGTTCAGTTCAGCCAGATGAGATCAGGAGGGTATGGATACACA 926  
QY 204 GlnValGlnAlaAspSerGlyAspGlyAspSerSerGluTyrGlnAspGlyLysGluPhe 223  
DB 927 CAGGTGATGACAGAGCATATATGAGACAGACACAGATATCAGATGATTAAGAATTT 986  
QY 224 GlyIleGlyAspLeuValTyrGlyLysIleLysGlyPheSerTyrTrpProAlaMetVal 243  
DB 987 GGAATAGGTGACCTCGTGGGGAAAGATCAAGGGCTTCTCTGCTGGCTGCCATGGTG 1046  
QY 244 ValSerTyrPheAlaThrSerLysArgGlnAlaMetSerGlyMetArgTyrValGlnTrp 263  
DB 1047 GTGCTCGAAGAACCCACTCCAGGACAGGCCATGCCGGAATGCGTGGATCACTGG 1106  
QY 264 PheGlyAspGlyLysPheSerGlyValSerAlaAspLysLeuValAlaLeuGlyLeuPhe 283  
DB 1107 TTTGGTGAATGCAAGTTTCTGAGATCTCTGTCGACAACTGGTGGCTTGGGGCTGTTT 1166  
QY 284 SerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLysAlaMetTyr 303  
DB 1167 AGCCAGCACTTATATCGGCTACCTTCATTAAGCTGTTTCTTAAGAAAGCCATGATAC 1226  
QY 304 HisAlaLeuGlnLysAlaArgValArgAlaGlyLysThrPheProSerSerProGlyAsp 323  
DB 1227 CACACTCGAAGAAAGCCAGGGTTCGAGCTGCGACAGACCTTCTCCAGCAGTCTGGAGAG 1286  
QY 324 SerLeuGluAspGlnLeuLysProMetLeuGluTyrAlaHisGlyLysPheLysProThr 343  
DB 1287 TCACTCGAGAGCACAGCGAAGCCATCGTGAAGTGGGCCCAAGGCTTCAAGCCACT 1346  
QY 344 GlyIleGlnGlyLysLysProAsnAsnThrGlnProValAlaAsnLysSerLysValArg 363  
DB 1347 GGGATCGAGGGCTCAAAACCAACAAGAGCAACCAAGCTTAATAGTCGAAGGTGGT 1406  
QY 364 ArgAlaGlySerArgLysLeuGluSerArgLysTyrGluAsnLysThrArgArgThr 383  
DB 407 CGTTCAGACAGTGAAGAACTTGAACCCAGAGAGCCGAGAAACAAGTGAAGACCGACA 1466  
QY 384 AlaAspAspSerAlaThrSerAspTyrCysProAlaProLysArgLeuLysThrAsnCys 403  
DB 1467 ACCAATGACTCGCTGCTTGAG--TCGCCCCACCCCAAGCGCTCAAGACAAATAGC 1523  
QY 404 TyrAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGluGlnMetAlaSerAsp 423  
DB 1524 TAT---GGCGGGAAGAGCCGAGGAGAGATGAGAGAGCCGAGAACGATGGCTTCGAA 1580  
QY 424 ValAlaAsnAsnLysSerSerLeuGluAspGlyCysLeuSerCysGlyArgLysAsnPro 443  
DB 1581 GTCAACCAACAACAGGGCAATCTGAGAACCGCTGTTTGTCTGTGGAAGAAGAACCTT 1640  
QY 444 ValSerPheHisPheProLeuPheGlnGlyLysLeuCysGlnThrCysArgAspArgPheLeu 463  
DB 1641 GTGTCTCTCCACCCCTCTTGAAGGATGGGCTCTGTGAGAGTTGCCGGATTCGTTCTTA 1700  
QY 464 GluLeuPheTyrMetTyrAspAspAspGlyTyrGlnSerTyrCysThrValCysCysGlu 483  
DB 1701 GAGCTCTTACATGATGATGAGAGCGGCTATCAGTCTACTGACCGGTGTGTGTGAG 1760  
QY 484 GlyArgGlnLeuLeuLeuCysSerAsnThrSerTyrCysArgCysPheCysValGluCys 503  
DB 1761 GGCCCTGAAGTCTGCTGTGAGTAACACAGAGCTGTCAGATGCTTCTGTGTGAGAGGT 1820  
QY 504 LeuGluValLeuValGlyThrGlyThrAlaAlaGluAlaLysLeuGlnGluProTrpSer 523

DB 1821 CTGAGAGTCTGCTGGGGCGAGGACACACTGAGATGCCAAGCTCCAGAAACCTTGAGC 1880  
QY 524 CysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgLysAspTyrAsn 543  
DB 1881 TGCTATATGTGCTCTCTCCAGCGGCTGCCATGGGGTCTCTCCAGCGAGGAAGAATTGGAAC 1940  
QY 544 ValArgLeuGlnAlaPhePheThrSerAspThrGlyLeu---GluTyrGlnAlaProLys 562  
DB 1941 ATGGGCTCGAAGACTTCTTCACTACTGATCTGACCTGGAAAGATTTGAGCAACCAAG 2000  
QY 563 LeuTyrProAlaIleProAlaAlaArgArgProIleArgValLeuSerLeuPheAsp 582  
DB 2001 TTGTATCCAGCAATTCCTGACAGCCAAAGAGGCCATTTAGAGTCTGCTCTGTTGAT 2060  
QY 583 GlyIleAlaThrGlyTyrLeuValLeuLysGluLeuGlyIleLysValGlyLysTyrVal 602  
DB 2061 GGAATGCAAGCGGCTACTGCTGCTCAAGAGTTGGGTATTAAGTGAAGAAAGTACATT 2120  
QY 603 AlaSerGluValCysGlnGluSerIleAlaValGlyThrValLysHisGlnGlyAsnIle 622  
DB 2121 GCTCCGAAGTCTGTGAGAGTCCATCCCTGTGGAACTGTTAAGCATGAAGCCAGATC 2180  
QY 623 LysTyrValAsnAspValArgAsnIleThrLysLysAsnIleGlnGluTyrGlyProPhe 642  
DB 2181 AATATATGCAATGACGTCCGAAATACCAAGAAATAATTGAAGAGTGGGCCCTTC 2240  
QY 643 AspLeuValIleGlyLysProCysAsnAspLeuSerAsnValAsnProAlaArgLys 662  
DB 2241 GACTTGATGATGGTGAAGCCATGCAATGATCTCTTAAGCTCAATCTGCCCGCAAA 2300  
QY 663 GlyLeuTyrGlnGlyThrGlyArgLeuPheGlnPheTyrHisLysLeuAsnTyrSer 682  
DB 2301 GGTATATGAGGAGCAGAGAGGCTCTTTCGAGTTTACACTTGTGTAATTATACC 2360  
QY 683 ArgProLysGlnGlyAspAspArgProPhePheTrpMetPheGluAsnValAlaMet 702  
DB 2361 CGCCCAAGAGAGGAGCAACCGTCAATCTTCTGATGTTGCAAAATGTTGGCCATG 2420  
QY 703 LysValGlyAspLysArgAspIleSerArgPheLeuGlnCysAsnProValMetIleAsp 722  
DB 2421 AAGTGAATGACAAAGAAAGCATCTCAAGATTCCTGGCAGTAAACCCAGTATGATGAT 2480  
QY 723 AlaIleLysValSerAlaAlaHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMet 742  
DB 2481 GCCATCAAGTGTCTGCTCTCAAGAGGCCGCTACTTCTGGGTAACCTACCCGGAATG 2540  
QY 743 AsnArgProValIleAlaSerLysAsnAspLysLeuGlnLeuGlnAspCysLeuGluTyr 762  
DB 2541 AACAGGCGCGTATGCTTCAAAAGATGATTAAGCTGAGCTGAGAGTGCCTGGAAGTTC 2600  
QY 763 AsnArgIleAlaLysLeuLysLysValGlnThrIleThrThrLysSerAsnSerIleLys 782  
DB 2601 AGTAGGACAGCAAAAGTTAAAGAAAGTGCAGACATTAACCAAGTCCGAATCCATCAGA 2660  
QY 783 GlnGlyLysAsnGlnLeuPheProValAlaMetAsnGlyLysGluAspValLeuTyrCys 802  
DB 2661 CAGGCAAAAACAGGCTTTCCTGTAGTCAAGATGCAAGAGCAGCCTTGTGTGTGC 2720  
QY 803 ThrGluLeuGlnArgIlePheGlyPheProValHisTyrThrAspValSerAsnMetGly 822  
DB 2721 ACTGAGCTCGAAAGATCTTCGCGTTCCTGCTCACTACCGAGCGTGTCCAATAGGGC 2780  
QY 823 ArgGlyAlaArgGlnLysLeuGlnGlyArgSerTyrSerValProValIleArgHisLeu 842  
DB 2781 CGCGCGCGCGTCAGAGAGTGTGGGAGGTCTCTGAGTGAACCGTCACTACAGACACTG 2840  
QY 843 PheAlaProLeuLysAspTyrPheAlaCysGlu 853  
DB 2841 TTTGCCCTTGAAGGACTACTTTCCTGTGAA 2873  
RESULT 7  
AF151969

**LOCUS** AF151969 4338 bp mRNA linear ROD 08-JUN-2000  
**DEFINITION** Mus musculus DNA cytosine-specific methyltransferase isoform 1  
**ACCESSION** AF151969  
**VERSION** AF151969.1 GI:8347117  
**KEYWORDS**  
**SOURCE** Mus musculus.  
**ORGANISM** Mus musculus  
**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
**AUTHORS** Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and Shen, Y.  
**TITLE** Cloning of full-length Dnmt3b cDNA and its alternative splicing isoforms in mouse embryonic tissue  
**JOURNAL** Unpublished  
**REFERENCE** 2 (bases 1 to 4338)  
**AUTHORS** Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and Shen, Y.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (17-MAY-1999) Department of Biochemistry and Molecular Biology, Institution of Basic Medical Sciences, Chinese Academy of Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005, P.R.China  
**FEATURES**  
**source**  
 1..4338  
 /organism="Mus musculus"  
 /strain="KM"  
 /db\_xref="taxon:10090"  
 /dev\_stage="8-9 day old embryo"  
 1..4338  
 /gene="Dnmt3b"  
 /note="Dnmt3b"  
 /function="methylates cytosine in DNA"  
 /note="alternatively spliced"  
 /codon\_start=1  
 /product="DNA cytosine-specific methyltransferase isoform 1"  
 /protein\_id="AAF74515.1"  
 /db\_xref="GI:8347118"  
 /translation="MKGDSRLNEERGASGYECIIIVNGNFSDDSDTKDAPSPVLE  
 ALCPEVCTPEGRSSRLSKREYSSLLNVTDMTGDGDDDDVDGSDILMPK  
 LITREYKTRTSFAVTRHNGTSLSERQASPRITRGQRHHVQEPVPEPATR  
 SRRRRSSASTPSSPASVDEMEVTPKSVSTPSVDSQDQDQEGMDITQVDAESIY  
 GDSTYQDDKEFGIDLVMGKIGFSWPMVSVKATSKRQAMPGRMVWQVMDGKPF  
 SEISADKLVALGFSOHFNLATFNKLVSKYKAMHYHLEKARVAGKTFSSPGESLED  
 QLKPMLEWAHGGFKPTGIEGLKPNKVPVVKSKVRRSDSRNLEPRRNRKSRRTTN  
 DSAASEPPPKRLKTNISYGGKDRGEDBSRMASEVTNNKNELDRLCSCKKNPV  
 FPHLVEGLCQSDRDFLELFYMDYEDGYQSYCTVCEGRELLCSNCSRCFCVFC  
 LEVLVAGTAEDAKLQEPMSCYMCLPQCHGVLRKRDWNRLLQDPTTDPDLREPER  
 EKLVPAPAAKRRPIRLVSLFDGIATGLVLKELGIKVKYKIASVECAESIAGTVKH  
 EQIKYVNDVRKITKNIEWGFDFLVIQSPCNDSNVNPARGLYEGTGRLLFEFY  
 HLLNTRPKGDNRPFFWMFENVAMKNDKDISRLACNPVMDIAIKVSAAHRY  
 FGNLPGMNRPMVASKNDKLEQDCLFERTAKLKVQTITTKSNSIRQGNQLFPVY  
 MNGKDDVLWCTELERIFGPFPAHYTDVSNMGRGARQKLLGRSWSVPVIRHLFAPLKDYF  
 ACE"  
**BASE COUNT** 1073 a 1124 c 1173 g 968 t  
**ORIGIN**  
**Alignment Scores:**  
 Pred. No.: 1.73e-190 Length: 4338  
 Score: 3748.00 Matches: 716  
 Percent Similarity: 88.52% Conservative: 55  
 Best Local Similarity: 82.20% Mismatches: 70  
 Query Match: 82.08% Indels: 30  
 DB: 10 Gaps: 10  
 US-09-720-086-8 (1-853) x AF151969 (1-4338)  
 Qy 1 MetLysGlyAspThrArgHisLeuAenGlyGluGluaspAlaGlyArgGluaspSer 20  
 Db 412 ATGAAGGGAGACAGACATCTGAATGAAGAAGAGGGTCCACGGGTATGAGGAGTGC 471

Qy 21 IleLeuValasnGlyAlaCysSerAspGlnSerSerAsp-----SerPro 35  
 Db 472 ATTATCGTTAATGGAACTTCAGTGACCACTCTCAGACAGAGGATGCTCCCTCACCC 531  
 Qy 36 ProIleLeuGluAlaIle-----ArgThrProGluIleArgGlyArgG 50  
 Db 532 CCGATCTTGAGGCAATCTGCACAGAGCCAGTCTGCACACAGACAGAGCCGCGG 591  
 Qy 51 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuSerTyThrGlnAspLeu 70  
 Db 592 TCAAGTCTCCGGGTGTCTAAGAGGGAGGTCTCCAGCCTTCTCAATATACAGCAGACATG 651  
 Qy 71 ThrGlyAspGlyAspGly-----GluAspGlyAspGlySerAspThrProVal 86  
 Db 652 ACAGGAGATGGACACAGAGATGATGAAGTAGATGATGGATGGCTCTGATATT---CTA 708  
 Qy 87 MetProLysLeuPheArgGlu-----ThrArgThrArgSerGluSerProAlaVal 103  
 Db 709 ATGCCAAAGCTCACCCGTGTGAGACCAAGGACACAGGACGCGCTCTGAAAGCCGCGTGC 768  
 Qy 104 ArgThrArgAsnAsnSerValSerSerArgGluArgHisArgProSerProArgSer 123  
 Db 769 CGAACCAGCATAGCAATGGGACCTCCAGCTTGAGAGGCAAGAGCCTCCCCAGAAATC 828  
 Qy 124 ThrArgGlyArgGlnGlyArgAsnHisValaspGluSerProValGluPheProAlaThr 143  
 Db 829 ACCCGAGGTGGCAGGCGCCGCCACCATGTGCAGAGTAGTACCTCTGGAGTTCCTCCGCTACC 888  
 Qy 144 ArgSerLeuArgArgAlaThrAlaSerAlaGlyThrProTyrProSerProSerProSer 163  
 Db 889 AGGTCTCGGAGACGTCGAGCATCGTCTTCAGCAAGACAGCCCATGTCATCCCTGCCAGC 948  
 Qy 164 SerTyrLeuThrIleAspLeuThrAspAspThrGluaspThrHisGlyThrProGlnSer 183  
 Db 949 -----GTGACTTTCATGGAAGAAGTG-----ACACTAAGAGC 981  
 Qy 184 SerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyGlyMetGluSerPro 203  
 Db 982 GTCAGTACCCCATCAGTTGACTTTGAGCCAGATGAGGAGGTATGATACCACA 1041  
 Qy 204 GlnValGluAlaAspSerGlyAspGlyAspSerSerGluTyrGlnAspGlyLysGluPhe 223  
 Db 1042 CAGTGGATGCAGAGACATATATGGAGACAGACAGATATCAGGATGATGAAGAGTIT 1101  
 Qy 224 GlyIleGlyAspLeuValThrGlyLysIleGlyPheSerTyrTrpProAlaMetVal 243  
 Db 1102 GGAATAGGTGACCTCGTGTGGGAAAGATCAAGGGCTTCTCTGTGGCTGCCATGGTG 1161  
 Qy 244 ValSerTrpLysAlaThrSerLysArgGlnAlaMetSerGlyMetArgTrpValGlnTrp 263  
 Db 1162 GTGTCTCGAAGCCACCTCCAGCGACAGCCATGCCGGAATGCCCTGGGTACAGTGG 1221  
 Qy 264 PheGlyAspGlyLysPheSerGluValSerAlaAspLysLeuValAlaLeuGlyLeuPhe 283  
 Db 1222 TTTGTGTCATGGCAAGTTCCTGAGATCTCTGACAACTGGTGGCTCTGGGGCTGTTTC 1281  
 Qy 284 SerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLysAlaMetTyr 303  
 Db 1282 AGCCAGCACTTTAATCTGGCTACCTTCAATAGTGGTTCCTTATAGGAGGCCATGTAC 1341  
 Qy 304 HisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSerSerProGlyAsp 323  
 Db 1342 CACACTCTGGAGAAAGCCAGGGTTCGAGCTGGCAAGACCTTCTCCAGCAGTCTCTGGAGAG 1401  
 Qy 324 SerLeuGluAspGlnLeuLysProMetLeuGluTrpAlaHisGlyGlyPheLysProThr 343  
 Db 1402 TCACCTGAGGACCCAGCTGAAGCCCATGCTGGAGTGGGCCACGGTGGCTTCAAGCTACT 1461  
 Qy 344 GlyIleGluGlyLysProAsnAsnThrGlnProValValAsnLysSerLysValArg 363  
 Db 1462 GGGATCGAGGGGCTCAAAACCCAAAGAAAGCAACAGTGGTGTAAATAGTCGAAGGTGCGT 1521





```
/function="de novo DNA methylation"  
/note="alternatively spliced product; contains Cys-rich  
region; C-terminal region is similar to corresponding  
region of DNA cytosine-5 methyltransferase 3A"  
/codon_start=1  
/product="DNA cytosine-5 methyltransferase 3B2"  
/protein_id="AAC40179.2"  
/db_xref="GI:6449472"  
/translation="MKGDSRLHNEEGAGSYECIIVNGNFSQSDTKADPPVLE  
AICPEVCTPSTRSSRLSKRGTSLLNYTQDMTGDGRDDDEVDDGNGSDILMPK  
LRETRKTRTRSPAVRTHSHGTSLSLRQASPRITRGRQGRHHVQEPYVFPATR  
SRRRRASASTPMSPASVDFMEETPKSVSTPSVDLSQDQEGMDTQVDAESRD  
GSTSEYQDKEFGIGDLVNGKIKGFSWPMVVMKATSKRQAMPGRWQWFGDGKF  
SEISADKLVALGLFSOHFNLATFNKLSYKAMHYHTLEKARVAGTKTFSSSPGSLSD  
OLKPMLEWAGGPKPTGLEGLPNKKQPKNSRRRTNDSSAESPPKRLKNTSYGG  
KRGDEESRERMASEVTNNKGNLEDRCLSCGKKNPUSPHPLREGGLCQSCRDFLL  
FYMDYDGYSTCTVCCBEGRELLCSNCCRCFCVECLVLAGAGTADAKQEPMS  
CYMGLPQRCGVLRRRKDMNRLQDFFTDPDLSEFEPKLYPAIPAKRRPRVLSL  
FDGIATGYLVLEKELIKVEKYIASEVCAESIAVGVKHEGQIKYVNDVRKITKNIEE  
WGPFDLVIGGSPCNOLSNVNPARKGLYEGRLLFFEFYLLNTRPKGDNRPFFWMP  
ENVVAMKVNDDKIIRFLACNPVMDAIKVSAAHRAHYFWGNLPGMNRPMVMAKNDKL  
ELQDCLEFSRTAKLKVOTITTKSNSIROGKNQLPVVMNGKDDVLWCTELERIFGP  
AHYTDVSNMGRGAROKLGRSVVPVIRHLFAPLKDYPFACE"
```

BASE COUNT 1043 a 1048 c 1106 g 938 t

ORIGIN

## Alignment Scores:

```
Pred. No.: 1.25e-186 Length: 4135  
Score: 3675.00 Matches: 703  
Percent Similarity: 86.80% Conservative: 53  
Best Local Similarity: 80.71% Mismatches: 65  
Query Match: 80.49% Indels: 50  
DB: 10 Gaps: 11
```

US-09-720-086-8 (1-853) x AF068627 (1-4135)

```
QY 1 MetLysGlyAspThrArgHisLeuAenGlyGluGluAspAlaGlyGlyArgGluAspSer 20  
DB 269 ATGAAGGAGAGACAGACATCTGAATGAAGAAGAGGGTCCAGCGGGTATGAGGAGTGC 328  
QY 21 IleLeuValAsnGlyAlaCysSerAspGlnSerSerAsp-----SerPro 35  
DB 329 ATTATCGTTAATGGGAACCTTCAGTGACCATCTCCAGACGACGAGGATGCTCCCTCACC 388  
QY 36 ProLeuLeuGluAlaIle-----ArgThrProGluIleArgGlyArgArg 50  
DB 389 CCAGTCTTGGAGGCAATTCGCACAGCCAGTCTGCACACAGAGACACAGAGGCCGAGG 448  
QY 51 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuSerTyrThrGlnAspLeu 70  
DB 449 TCAAGCTCCCGCTGCTTAAGAGGAGGTCTCCAGGCTTCTGAATTACACGACGACATG 508  
QY 71 ThrGlyAspGlyAspGly-----GluAspGlyAspGlySerAspThrProVal 86  
DB 509 ACAGAGATGGAGACAGATGATGAATAGTAGATGGGAATGGCTCTGATATT---CTA 565  
QY 87 MetProLysLeuPheArgGlu-----ThrArgThrArgSerGluSerProAlaVal 103  
DB 566 ATGCCAAAGCTCACCGCTGAGACCAAGGACACCCAGGACGCGCTCTGAAGGCCCGCTGTC 625  
QY 104 ArgThrArgAenAsnAsnSerValSerSerArgGluArgHisArgProSerProArgSer 123  
DB 626 CGAACCCACATAGCAATGGGACCTTCAGCTTGGAGAGGCAAGAGGCTCCCCCAGAAATC 685  
QY 124 ThrArgGlyArgGlnGlyArgHisValAspGluSerProValGluPheProAlaThr 143  
DB 686 ACCCGAGGTCCGCGGCGGCCACCATGTGCAGAGGTACCTGTGGAGTTTCCGGCTACC 745  
QY 144 ArgSerLeuArgArgAlaThrAlaSerAlaGlyThrProThrProSerProSer 163  
DB 746 AGGTCTCGAGACGTCGACATCGTCTTCAGCAAGCAGCAGCATGGTTCATCCCTGCCAGC 805  
QY 164 SerTyrLeuThrIleAspLeuThrAspThrGluAspThrHisGlyThrProGlnSer 183
```

```
DB 806 -----GTCGACTTCATGGAAGAAGTG-----ACACCTAAGAGC 838  
QY 184 SerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyMetGluSerPro 203  
DB 839 GTCAGTACCCCATCAGTTGACTTGAGCCAGGATGAGATCAGAGGGTATGATACCACA 898  
QY 204 GlnValGluAlaAspSerGlyAspGlyAspSerSerGluTyrGlnAspGlyLysGluPhe 223  
DB 899 CAGGTGGATGCAGACAGACAGATGGAGACAGACAGATATCAGGATGATAAAGAGTTT 958  
QY 224 GlyIleGlyAspLeuValTyrGlyLysIleLysGlyPheSerTyrTyrProAlaMetVal 243  
DB 959 GGAATAGGTGACCTCGTGTGGGAAAGATCAAGGGCTTCTCTGGTGGCTGCCCATGGTG 1018  
QY 244 ValSerTyrLysAlaThrSerLysArgGlnAlaMetSerGlyMetArgTyrValGlnTyr 263  
DB 1019 GTGTCTCGAAAGCCACCTCAAGCGACAGCCCATGCCCGGAATGCGCTGGGTACAGTGG 1078  
QY 264 PheGlyAspGlyLysPheSerGluValSerAlaAspLysLeuValAlaLeuGlyLeuPhe 283  
DB 1079 TTTGGTGATGGCAAGTTTCTGAGATCTCTGCTGACAAACTGGTGGCTCTGGGGCTGTT 1138  
QY 284 SerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLysAlaMetTyr 303  
DB 1139 AGCCAGCATTTTAATCTGGCTACTCTCAATAAGCTGGTTCTTATAGGAAGCCATGTAC 1198  
QY 304 HisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSerSerProGlyAsp 323  
DB 1199 CACACTCTGGAGAAAGCCAGGGTTCGAGCTGGCAAGACCTTCTCCAGCAGTCTCTGGAG 1258  
QY 324 SerLeuGluAspGlnLeuLysProMetLeuGluTyrAlaHisGlyGlyPheLysProThr 343  
DB 1259 TCACCTGGAGGACAGCTGAAGCCCATGCTGGAGTGGGCCACGCTTCAAGGCTACT 1318  
QY 344 GlyIleGluGlyLeuLysProAsnAsnThrGlnProValValAsnLysSerLysValArg 363  
DB 1319 GGGATCGAGGGCTCAAAACCCACAAAGAACCCACCA-----1354  
QY 364 ArgAlaGlySerArgLysLeuGluSerArgLysTyrGluAsnLysThrArgArgArgThr 383  
DB 1355 -----GAGAACAAAGTCGAAGACGCACA 1378  
QY 384 AlaAspAspSerAlaThrSerAspTyrCysProAlaProLysArgLeuLysThrAsnCys 403  
DB 1379 ACCAATGACTCTCTGCTTCTGAG---TCCCCCCCCCAAGCGCTCAAGACAATAGC 1435  
QY 404 TyrAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGluGlnMetAlaSerAsp 423  
DB 1436 TAT---GGCGGGAAGGACCCGAGGGGAGGATGAGGAGCCGAGAACGGATGGCTTCTGAA 1492  
QY 424 ValAlaAsnAsnLysSerSerLeuGluAspGlyCysLeuSerCysGlyArgLysAsnPro 443  
DB 1493 GTCACCAACAAAGGGCAATCTGGAAGACCGCTGTTTGTCTGTGGAAAGAAAGAACCTT 1552  
QY 444 ValSerPheHisProLeuPheGluGlyGlyLeuCysGlnThrCysArgAspArgPheLeu 463  
DB 1553 GTGTCTCTCCACCCCTCTTTGAGGGTGGGCTCTGTGAGAGTTCGCGGATCGCTTCTCTA 1612  
QY 464 GluLeuPheTyrMetTyrAspAspGlyTyrGlnSerTyrCysThrValCysCysGlu 483  
DB 1613 GAGCTCTTCTACATGATGATGAGGCGGTATCAGTCTTACTGACCGCTGTGTGTGTAG 1672  
QY 484 GlyArgGluLeuLeuLeuCysSerAsnThrSerCysCysArgCysPheCysValGluCys 503  
DB 1673 GGCCGTAAGTGTGCTGTGTCAGTAACACAGCTGCTGTCAGATGCTTCTGTGTGAGTGT 1732  
QY 504 LeuGluValLeuValGlyThrGlyThrAlaAlaGluAlaLysLeuGlnGluProThrSer 523  
DB 1733 CTGGAGGTGCTGGTGGCGGACAGCTGAGGATGCAAGCTGCAGAGCCCTGGAGC 1792  
QY 524 CysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgLysAspThrAsn 543
```

Dh 1793 TGCTATATATGCTCCCTCAGCGCTGCATGGGGTCTCCGACGACGAGAAAGATGGAAC 1852  
Qy 544 ValArgLeuGlnAlaPhePheThrSerAspThrGlyLeu---GluTyrGlnAlaProLys 562  
Db 1853 ATGCGCTGCAAGACTCTTCACACTGATCTCGACTGGAAGAAATTTAGACCAACCAAG 1912  
Qy 563 LeuTyrProAlaIleProAlaAlaArgArgArgProIleArgValLeuSerLeuPheAsp 582  
Db 1913 TTGTACCCAGCAATCTCTGACGCCAAAGAGGCCCATTAAGTCTCTGTTGAT 1972  
Qy 583 GAllyleAlaThrGlyTyrLeuValLeuLysGluLeuGlyIleLysValGlyLysTyrVal 602  
Db 1973 GGAATTGCAACGGGGTACTGCTGCTCAAGAGAGTTGGTATTAAGTGAAGAAAGTACATT 2032  
Qy 603 AlAserGluValCyseGluGluSerIleAlaValGlyThrValLysHisGluGluLysHis 622  
Db 2033 GCCTCCGAAGTCTGTTCAGAGTCCATCGCTGTGGAACTGTTAAAGCATGAAGCCAGATC 2092  
Qy 623 LysTyrValAsnAspValArgAsnIleThrLysLysAsnIleGluGluTTPGlyProPhe 642  
Db 2093 AAATATGTCATGACGCTCCGAAATACCAAGAAATATTAAGAGTGGGGCCCGCTTC 2152  
Qy 643 AspLeuValIleGlyLysSerProCyseAsnAspLeuSerAsnValAsnProAlaArgLys 662  
Db 2153 GACTTGGTATGTGTGGAGCCATGATGATCTCTTAACGTCAATCTGCCCGCAAA 2212  
Qy 663 GlyLeuTyrGluGlyThrGlyArgLeuPheGluPheGlyThrHisLeuLeuAsnTyrSer 682  
Db 2213 GGTTATATGAGGGGCAGAGAGGCTCTTTCAGATTTCACACTGCTGAATTATATACC 2272  
Qy 683 ArgProLysGluGlyAspAspArgProPhePheThrMetPheGluAsnValValAlaMet 702  
Db 2273 CGCCCCAAGAGAGGCGCAACAGCTCATCTTCTGATGTTTCAGAAATGTGGCCATG 2332  
Qy 703 LysValGlyAspLysArgAspIleSerArgPheLeuGluCyseAsnProValMetIleAsp 722  
Db 2333 AAATGTAATGACAAAGAAAGATCTCAAGATTCCTGCATGTAAACCAAGATGATGAT 2392  
Qy 723 AlaIleLysValSerIleAlaHisArgAlaArgTyrPheTTPGlyAsnLeuProGlyMet 742  
Db 2393 GCCATCAAGGTGTCTGTCTCACAGGGCCCGGTACTTCTGGGTAACTTACCCTGGAATG 2452  
Qy 743 AsnArgProValIleAlaSerLysAsnAspLysLeuGluLeuGlnAspCyseLeuGluTyr 762  
Db 2453 AACAGGCCCGTGTGCTTCAAGATGATTAAGCTCGAGCTGCAGAGCTCCGTGAGATTC 2512  
Qy 763 AsnArgIleAlaLysLeuLysValGlnThrIleThrLysSerAsnSerIleLys 782  
Db 2513 AGTAGGACGCAAAAGTTAAAGAAAGTGCAGACATTAACCAAGCTCGAATCCATCAGA 2572  
Qy 783 GlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysGluAspValLeuTTPCyse 802  
Db 2573 CAGGGCAAAACCAAGCTTTTCCGTGATGATGATGATGATGATGATGATGATGATGATG 2632  
Qy 803 ThrGluLeuGlnArgIlePheGlyPheProValHisTyrThrAspValSerAsnMetGly 822  
Db 2633 ACTGAGCTCGAAGGATCTTGGCTTCCCTGCTCAGACGAGCTGTCACCATGATGATGATG 2692  
Qy 823 ArgGlyAlaArgGlnLysLeuLeuGlyArgSerTTPSerValProValIleArgHisLeu 842  
Db 2693 CGCGCGCGCGCTGCAAGAGCTCTGCGCAGGCTCTGAGTGAACCGGTCAATCAGACACTG 2752  
Qy 843 PheAlaProLeuLysAspTyrPheAlaCyseGlu 853  
Db 2753 TTTGCCCTTGAAGGACTACTTGTCTGTGAA 2785  
RESULT 9 AF151974 4163 bp mRNA linear ROD 08-JUN-2000  
AF151974  
DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 6  
(Dnmt3b) mRNA, complete cds.  
ACCESSION AF151974  
VERSION AF151974.1 GI:8347130

KEYWORDS  
SOURCE Mus musculus.  
ORGANISM Mus musculus.  
REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 4163)  
Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and  
Shen,Y.  
TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing  
isoforms in mouse embryonic tissue  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 4163)  
Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and  
Shen,Y.  
TITLE Direct Submission  
JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular  
Biology, Institution of Basic Medical Sciences, Chinese Academy of  
Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,  
P.R.China  
FEATURES  
source location/qualifiers  
1..4163  
/organism="Mus musculus"  
/strain="KM"  
/db\_xref="taxon:10090"  
/dev\_stage="8-9 day old embryo"  
1..4163  
/gene="Dnmt3b"  
297..2816  
/gene="Dnmt3b"  
/function="methylates cytosine in DNA"  
/note="alternatively spliced"  
/codon\_start=1  
/product="DNA cytosine-specific methyltransferase isoform  
6"  
/protein\_id="AA174520.1"  
/db\_xref="GI:8347131"  
/translation="MKDSDSRHLNBEAGAGYEBCIIVNGNFSDDSDTKDAPSPVLE  
AICTEPVCTPETGRSSRLSKREYSLNLTQDMTGDGDDDEVDGSGDILMPK  
LIREKDTPTSRSPASVATRTSHSGTSLRROPASPRITRGRGRHGVQEVPERPATR  
SRRRASGSASTPWSSPASVPMEEVTPSPVSTPVSIDSQDQDEMDTTQVDNESTY  
GDSTEVDDKEFGIDLVWGKIKGSPWPAWVWSKAIKROAMPQWQWQFDDGKF  
SEISADKVALGFSQHNFLATFNKLVSRKAMHTLEKARVAGATFSSPGSLED  
OLKPMLEMAHGKRPTEGELKPNKKOPRNKSRRTTNDASASESPPKRLKTSYGG  
KRGDEESRERNASEVTNKNLGNLEDRCLSGCKNPVSFHLPEGLCGSCRRFLEL  
FMYDEDDGYOSTVCEGRILLCSNTSCRCCEVECLELVAGTAEPAKLOEPMWS  
CYMLDPRGCHVRRKDMNRLQDFTDDEFEPEPKLYPAIPAARKRPVLSL  
FDGATGVYLVKEGLIKVEKYLASEVCASIAVGTVHGOIKYVADVRITTKNIEE  
WGFPIVLVIGSPCNLDLNVNPAKGLIEGTRLEFEPFHLIATYTRKEGNNRPFMW  
ENVMKVKNDKDIISRELAENPVMIDAIKVSAAHARARYFWGNLPGNNRPVMAKNDKL  
ELQDCLEPSTAKLKVKQVITTKSNSIRQGNQLEFPVVMNGKDDVLWCTELERIFGFP  
AHYTVSNNGRGROROKLIGRGSVPVIRHLFAPLKPDPACE"  
BASE COUNT 1032 a 1074 c 1128 g 929 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 4,85e-186 Length: 4163  
Score: 3664.00 Matches: 702  
Percent Similarity: 86.68% Conservative: 53  
Best Local Similarity: 80.60% Mismatches: 66  
Query Match: 80.25% Indels: 50  
DB: 10 Gaps: 11  
US-09-720-086-8 (1-853) x AF151974 (1-4163)  
Qy 1 MetLysGlyAspThrArgHisLeuAsnGlyGluGluAspAlaGlyGlyArgGluAspSer 20  
Db 297 ATGAAGGAGACAGACATCTGATGAAGAAGAGGTGCAGCGGATGATGAGAGAGTGC 356  
Qy 21 IleLeuValAsnGlyAlaCyseSerAspGlnSerSerAsp-----SerPro 35  
Db 357 ATTAATCGTTAATGGGAATTCAGTGACAGAGTCTCAGACACGAAGAGATGCTCCCTCACCC 416  
Qy 36 ProIleLeuGlnAlaIle-----ArgThrProGluIleArgGlyArgArg 50

Db 417 CCAGTCTTGGAGGCAATCTGCACAGAGCCAGTCTGCACACAGAGCCAGAGGCGCGCAGG 476  
Qy 51 SerSerArgLeuSerLysArgGluValSerSerLeuLeuSerTyrThrGlnAspLeu 70  
Db 477 TCAGCTCCCGGCTGTCTAAGAGGGAGGCTCCAGCCTTCTGAATTACACGACGACATG 536  
Qy 71 ThrGlyAspGlyAspGly-----GluAspGlyAspGlySerAspThrProVal 86  
Db 537 ACAGGAGATGGAGACAGAGATGATGAAGTAGATGATGGGAATGGCTCTGATATT--CTA 593  
Qy 87 MetProLysLeuPheArgGlu-----ThrArgThrArgSerGluSerProAlaVal 103  
Db 594 ATGCCAAGACTCACCGTGGAGCAAGGACACACGAGGACGCTCTGAAGAGCCCGGCTGTC 653  
Qy 104 ArgThrArgAsnAsnSerValSerSerArgGluArgHisArgProSerProArgSer 123  
Db 654 CGAACCCGACATAGCAATGGGACCTCCAGCTTGAGAGGCAAGAGGCTCCCGCCAGAAATC 713  
Qy 124 ThrArgGlyArgGlnGlyArgAsnHisValAspGluSerProValGluPheProAlaThr 143  
Db 714 ACCCGAGGTGGCGCGGCCACCATGTGCAGAGTACCTGTGGAGTTTCCGGCTACC 773  
Qy 144 ArgSerLeuArgArgAlaThrAlaSerAlaGlyThrProTrpProSerProSer 163  
Db 774 AGGTCTCGAGACGTGAGCATCTGCTTCAGCAAGACAGCGCATGTGTCATCCCTGCCAGC 833  
Qy 164 SerTyrLeuThrIleAspLeuThrAspAspThrGluAspThrHisGlyThrProGlnSer 183  
Db 834 -----GTGCACCTTCATGGAAGAAGTG-----ACACCTTAAGAGC 866  
Qy 184 SerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyGlyMetGluSerPro 203  
Db 867 GTCAGTACCCCATCAGTTGACTTGCAGCAGATGGAGTCAGGAGGTATGGATACACACA 926  
Qy 204 GlnValGluAlaAspSerGlyAspGlyAspSerSerGluTyrGlnAspGlyLysGluPhe 223  
Db 927 CAGGTGGATGCAGAGAGCATATATGAGACAGACAGCATGATCAGAGATGATAAAGAGTTT 986  
Qy 224 GlyIleGlyAspLeuValTrpGlyLysIleLysGlyPheSerTrpTrpProAlaMetVal 243  
Db 987 GGNATAGTGTACCTGTGTGGGAAAGATCAAGGGCTTCTCGTGGCTGCCATGGTG 1046  
Qy 244 ValSerTrpLysAlaThrSerLysArgGlnAlaMetSerGlyMetArgTrpValGlnTrp 263  
Db 1047 GTGTCTCGAAGCCACCTCCAAAGCGACAGGCCATGCCGGAATGCCGTGGGTACAGTGG 1106  
Qy 264 PheGlyAspGlyLysPheSerGluValSerAlaAspLysLeuValAlaLeuGlyLeuPhe 283  
Db 1107 TTTGGTGTGGCAAGTTTCTGAGATCTCTGCTGACAACTGGTGGCTCTGGGCTGTTC 1166  
Qy 284 SerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLysAlaMetTyr 303  
Db 1167 AGCCAGCACTTTAAATCTGGCTACCTTCAATTAAGCTGGTTTCTTATAGGAAGGCCATGTAC 1226  
Qy 304 HisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSerSerProGlyAsp 323  
Db 1227 CACACTCTGGAGAAAGCCAGGTTTCGAGCTGGCAAGACCTTCTCCAGCAGTCTCTGGAGAG 1286  
Qy 324 SerLeuGluAspGlnLeuLysProMetLeuGluTrpAlaHisGlyGlyPheLysProThr 343  
Db 1287 TCACGTGAGGACCACTGAAGCCCATCTGGAGTGGCCACCGTGGCTTCAAGCCTACT 1346  
Qy 344 GlyIleGluGlyLeuLysProAsnAsnThrGlnProValValAsnLysSerLysValArg 363  
Db 1347 GGGATCGAGGGGCTCAAAACCAACAAAGAACACCC----- 1382  
Qy 364 ArgAlaGlySerArgLysLeuGluSerArgLysTyrGluAsnLysThrArgArgThr 383  
Db 1383 -----GAGAACAAAGTCCGAAGACGCACA 1406  
Qy 384 AlaAspSerAlaThrSerAspTyrCysProAlaProLysArgLeuLysThrAsnCys 403  
Db 403 ----- 476

Db 1407 ACCAATGACTCTCTCTCTCTGAG---TCCCCCCCAACCCCAAGCGCTCAAGACAAATAGC 1463  
Qy 404 TyrAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGluGlnMetAlaSerAsp 423  
Db 1464 TAT---GGCGGGAAGGACCGAGGGAGGATGAGGAGCCGAGAACCGATGCTTCTTGAA 1520  
Qy 424 ValAlaAsnAsnLysSerSerLeuGluAspGlyCysLeuSerCysGlyArgLysAsnPro 443  
Db 1521 GTCACCAACAACAAGGCAATCTGGAAGACCGCTGTGTCTCTGTGGAAGAAGAACCCCT 1580  
Qy 444 ValSerPheHisProLeuPheGluGlyGlyLeuCysGlnThrCysArgAspArgPheLeu 463  
Db 1581 GTGTCTCTCCACCCCTCTTTGAGGGTGGCTCTGTGAGAGTTGCCGGGATCGCTTCTTA 1640  
Qy 464 GluLeuPheTyrMetTyrAspAspAspGlyTyrGlnSerTyrCysThrValCysCysGlu 483  
Db 1641 GAGCTCTTCTACATGATGATGAGGACGGCTATCAGTCTCTACTGACCGTGTGTGTGAG 1700  
Qy 484 GlyArgGluLeuLeuCysSerAsnThrSerCysCysArgCysPheCysValGluCys 503  
Db 1701 GGCGGTGAACCTGTCTGTGAGTAAACACAGCTGTGTGAGATGCTTCTGTGTGGAGTCT 1760  
Qy 504 LeuGluValLeuValGlyThrGlyThrAlaAlaGluAlaLysLeuGlnGluProTrpSer 523  
Db 1761 CTGGAGGTGCTGTGGCGCAGCAGCTGAGATGCCAAGCTGCAGAACCCCTGGAGC 1820  
Qy 524 CysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgArgLysAspTrpAsn 543  
Db 1821 TGTATATGTGCTCCCTCAGCGTGCCTGGGGTCTCTCCGACGCGAGAAAGATTGGAAC 1880  
Qy 544 ValArgLeuGlnAlaPhePheThrSerAspThrGlyLeu---GluTyrGluAlaProLys 562  
Db 1881 ATGCGGCTGCAGACTTCTCACTACTGATCTCTACCTGGAAGAATTTGAGCCACCAAG 1940  
Qy 563 LeuTyrProAlaIleProAlaAlaArgArgArgProIleArgValLeuSerLeuPheAsp 582  
Db 1941 TTGTACCCAGCAATCTCTGAGCCAAAGAGGAGGCCATATAGAGTCTCTCTCTGTGTAT 2000  
Qy 583 GlyIleAlaThrGlyTyrLeuValLeuLysGluLeuGlyIleLysValGlyLysTyrVal 602  
Db 2001 GGAATTTGCAACGGGTACTTTGGTGTCAAGGAGTTGGGTATTAAAGTGGAAGATACAT 2060  
Qy 603 AlaSerGluValCysGluGluSerIleAlaValGlyThrValLysHisGluGlyAsnIle 622  
Db 2061 GCTTCGAGTGTGCGAGAGTCCATCGCTGTGGAACTGTTAAGCATGAAGGCCAGATC 2120  
Qy 623 LysTyrValAsnAspValArgAsnIleThrLysLysAsnIleGluGluTrpGlyProPhe 642  
Db 2121 AAATATGTCAATGACGTCCGAAAAATCACAAGAAAAAATATTGAAGAGTGGGGCCGTTTC 2180  
Qy 643 AspLeuValIleGlyGlySerProCysAsnAspLeuSerAsnValAsnProAlaArgLys 662  
Db 2181 GACTTGGTATTGTTGGAAGCCCATGCAATGATCTCTAAGCTCAATCCTCCCGCAAA 2240  
Qy 663 GlyLeuTyrGluGlyThrGlyArgLeuPhePheGluPheTyrHisLeuLeuAsnTyrSer 682  
Db 2241 GGTATTATGAGGGCACAGGAAGGCTCTTCTTCAGATTTTACCACCTTCTGAAATTATACC 2300  
Qy 683 ArgProLysGluGlyAspAspArgProPhePheTrpMetPheGluAsnValValAlaMet 702  
Db 2301 CGCCCCAAGGAGGCGCAACCGTCCATTCTTCTGGATGTTTCGAGAAATGTTGTGGCCATG 2360  
Qy 703 LysValGlyAspLysArgAspIleSerArgPheLeuGluCysAsnProValMetIleAsp 722  
Db 2361 AAGTGAATGACAAGAAAGACATCTCAAGATTCTTGGCATGTAACCCCAAGTATGATCGAT 2420  
Qy 723 AlaIleLysValSerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMet 742  
Db 2421 GCCATCAAGGTGTCTGTGCTCACAGGCGCGGTACTTCTGGGGTAACTACCCCGAATG 2480  
Qy 743 AsnArgProValIleAlaSerLysAsnAspLysLeuGluLeuGlnAspCysLeuGluTyr 762  
Db 2481 AACAGGCGCGTGTGCTTCAAGAAATGATAAGTTCGAGCTGCAGACTGCTGTGGAGTTC 2540

QY	763	AsnArgIleAlaIleuLeuValGlnThrIleThrThrIleValSerAsnSerIleIleVal	782
Db	2541	AGTGGACAGCAAGTTAAAGAAAGTGCAGACATTAACCAAGTCAGCTTCATCAGA	2600
QY	783	GlnGlyValSerAsnGlnLeuPheProValValMetAsnGlyValSerGluAspValIleuTrpCys	802
Db	2601	CAGGGCAAAACACAGCTTTTCCCTGTATCATGATGCGAAGGACGAGCTTTGTGTGC	2660
QY	803	ThrGluLeuGluArgIlePheGlyPheProValHisIleThrAspValSerAsnMetGly	822
Db	2661	ACTGAGCTCGAAGAGATCTTCGGCTTCCCTGCTCACTACACGAGCTGTCCACATGAGC	2720
QY	823	ArgGlyValAlaArgIleuLeuLeuGluValArgSerTrpSerValProValIleArgHisLeu	842
Db	2721	CGGGCGCGCCGTCAGAGCTGCGGCGAGGTCCTGGAAGTGAACCGGTATCAGACACCTG	2780
QY	43	PheAlaProLeuLeuAspTyrPheAlaCysGlu	853
Db	2781	TTTGCCCCCTTGAGAGACTACTTGTGCTGTGAA	2813
RESULT 10			
LOCUS	AF151970		
DEFINITION	AF151970	4278 bp	mRNA linear ROD 08-JUN-2000
ACCESSION	Mus musculus DNA cytosine-specific methyltransferase isoform 2		
VERSION	(Dnmt3b) mRNA, complete cds.		
KEYWORDS	AF151970		
SOURCE	AF151970.1 GI:8347119		
ORGANISM	Mus musculus.		
REFERENCE	Mus musculus		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclirognathi; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 4278)		
JOURNAL	Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and Shen,Y.		
AUTHORS	Cloning of full-length Dnmt3b cDNA and its alternative splicing isoforms in mouse embryonic tissue		
TITLE	2 (bases 1 to 4278)		
JOURNAL	Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and Shen,Y.		
AUTHORS	Direct Submission		
TITLE	Submitted (17-MAY-1999)		
JOURNAL	Department of Biochemistry and Molecular Biology, Institution of Basic Medical Sciences, Chinese Academy of Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005, P.R.China		
FEATURES			
source	Location/Qualifiers		
	1..4278		
	/organism="Mus musculus"		
	/strain="KM"		
	/db_xref="taxon:10090"		
	/dev_stage="8-9 day old embryo"		
	1..4278		
	/gene="Dnmt3b"		
	412..2931		
	/gene="Dnmt3b"		
	/function="methylates cytosine in DNA"		
	/note="alternatively spliced"		
	/codon_start=1		
	/product="DNA cytosine-specific methyltransferase isoform 2"		
	/protein_id="AAF74516.1"		
	/db_xref="GI:8347120"		
	/translation="WKGDGRHLNNEEGAGYEECTIVNGNFSDDSDTDAPAPVIE		
	LITETPVCPTETGRSSRLSKREVSLINLYQDMTGDGRDDEVDNDSILMPRL		
	LRTEKDTSTRSSPAVATIRNSNGTSLERQASPRITRGGRHLYOEYVPEPATRR		
	SRRRASSASTPWSPASVDPMETPTSPVSLDSOGDQSGMDTTQVDAESITY		
	GDSTYQDDKEFGGLDLWGKIKGFSWMPAMVVSWKATSKQAMPKRWQMGDGKR		
	SEISADLVALGIFSQFNLAIFNKLVSRKAWHTLEKARVRAQTFSSPESLSD		
	OLKPLMEAHAGFKPTGIEGLKPNKQPEKSRRTTNDASAESPPKRLKINSYGG		
	KDRGEESRERMASEVTNNKGNLDRCLSCGKKNPVSFHLFEGIGCSCRPLFEL		
	FMYIEDGIOSTYCTVCCBGRILLCLNSITSCRCFCVECLEVLVAGTAEDAKLOE		
	PWS		

BASE COUNT		1055 a	1113 c	1154 g	956 t
ORIGIN					
Alignment Scores:					
Pred. No.:	5, 01e-186	Length:	4278		
Score:	3664.00	Matches:	702		
Percent Similarity:	86.68%	Conservative:	53		
Best Local Similarity:	80.60%	Mismatches:	66		
Query Match:	80.25%	Gaps:	50		
DB:	10	Indels:	11		
US-09-720-086-8 (1-853) x AF151970 (1-4278)					
QY	1 MetHysGIAspThrArgHisLeuAsnGLyGIuAspAlaGLyIArgGIuAspSer	20			
DB	412 ATGAGAGGAGACAGCAGACACTCTGATATAGAGAGAGGCTCCACGGGATATAGAGACTGC	471			
QY	21 ILeuEuAlaSnGLyAlaCysSerAspGlnSerSerAsp-----SerPro	35			
DB	472 ATTATCGTTAATGGAACTTCAGTGAACCAAGTCTTCAGACACGAAGAGATGCTCCCTCAACC	531			
QY	36 ProfileuGIuAlaIle-----ArgThrProGluIleArgGIuArg	50			
DB	532 CCACTCTTGAGAGCAATCTGCACAGACCAGTCTGCACACACAGACCAAGAGCCCGCAGG	591			
QY	51 SerSerSerArgLeuSerIlyAspGIuAlaSerSerLeuSerIlyThrGlnAspLeu	70			
DB	592 TCAAGCTCCGGCTGCTTAAAGAGGAGGTCTTCAAGCTTCTGAATTACACGACGACATG	651			
QY	71 ThrGIuAspGIuAspGly-----GluAspGIuAspGlySerAspThrProAla	86			
DB	652 ACAGAGATGGAGACAGACAGATGATGAACTGATGAGGCAATGCTGTGATATT---CTA	708			
QY	87 MetProIlySleuPheArgGlu-----ThrArgThrArgSerGluSerProAlaVal	103			
DB	709 ATGCCAAGCTACCCGTGAGACCAAGSACACGAGACGCGCTGAAAGCCCGGCTGTC	768			
QY	104 ArgThrArgAspAsnAsnSerValSerSerIArgGluArgHisArgProSerProArgSer	123			
DB	769 CGAACCCCAATGACATGGAGACCTCCAGCTTGGAAAGGCAAAAGACCTCCCAAGATC	828			
QY	124 ThrArgGIuArgGlnIlyArgAsnHisValaAspGluSerProValGluPheProAlaThr	143			
DB	829 ACCGAGAGTCCGACGGGCCCCACACANTGTGACAGAGTACCCTGTGAGATTCCGCGTACC	888			
QY	144 ArgSerLeuArgArgArgAlaThrAlaSerAlaGLyIThrProThrProSerProProSer	163			
DB	889 AGGCTTCGAGAGCGTGCAGCATCGCTTCAGAAAGCAGCCATGATCAATCCCTCGCACGC	948			
QY	164 SerIlyrIleuThrIleAspLeuThrAspAspThrGluAspThrHisGLyIThrProGlnSer	183			
DB	949 -----GTCACTTCATGAGAAAGTG-----ACACCTAAGAGC	981			
QY	184 SerSerThrProThrIlaArgLeuAlaGlnAspSerGlnGlnGlyMetGluSerPro	203			
DB	982 GTCGATACCCCATCACTTGAATTCAGCCAGAGTGAAGATCAGAGAGGATGATATACACA	1041			
QY	204 GlnAlaGluAlaAspSerGlyAspGlyAspSerSerGluIlyrGlnAspGlyIlySGluPhe	223			
DB	1042 CAGGTGATGACAGAGCATATATGAGACACACAGAGATCAGGATGATAAAGAGTTT	1101			
QY	224 GlnIlyIleGIuAspLeuValTyrGlyIlyHisIleIlySGlyPheSerTyrProAlaMetVal	243			
DB	1102 GGAATTAGGTGACTCTGTGTGGGAAAGATCAAGGGCTTCTCTGTGTGCTCCATGAGTG	1161			
QY	244 ValSerTyrPylalThrSerIlyArgGlnAlaMetSerGlyMetArgTyrAlaGlnTyr	263			

1162	Db	GTGTCTCTGGAAAGCCACCTCTCCAGCGCACAGGCCATGCCCGGAATGCGCTGGGTACAGTGG	1221
264	Qy	PheGlyAspGlyLysPheSerGluValSerAlaAspLysLeuValAlaLeuGlyLeuPhe	283
1222	Db	TTTGGTGATGCAAGTTTCTCAGAGTCTCTGTGCAAAACTGCTGGCTCTCTGGGGCTGTTC	1281
284	Qy	SerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLysAlaMetTyr	303
1282	Db	AGCCAGCACTTTAATCTGGCTACCTTCAATAAGCTGGTTTCTTATAGGAAGGCCATGTAC	1341
304	Qy	HisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSerSerProGlyAsp	323
1342	Db	CACACTCTGGAGAAAGCCAGGGTTCTGAGCTGGCAAGACCTTCTCCAGCAGTCTCTGGAGAG	1401
324	Qy	SerLeuGluAspGlnLeuLysProMetLeuGluTTPAlaHisGlyGlyPheLysProThr	343
1402	Db	TCATCTGGAGGACCACTGAGCCCAATGCTGGAGTGGGCCCACGGTGGCTTCAAGCCCTACT	1461
344	Qy	GlyIleGluGlyLeuLysProAsnAsnThrGlnProValValAsnLysSerLysValArg	363
1462	Db	GGGATCGAGGGCCTCAAAACCCCAACAGGAAGCAACCA-----	1497
364	Qy	ArgAlaGlySerArgLysLeuGluSerArgLysTyrGluAsnLysThrArgArgArgThr	383
1498	Db	-----GAGAACAAAGTTCGAGACGCACAA	1521
384	Qy	AlaAspAspSerAlaThrSerAspTyrCysProAlaProLysArgLeuLysThrAsnCys	403
1522	Db	ACCAATGACTCTGTCTTCTTGAG--TCCCCGCCCAACCCAGCGCTCAAGACAAATAGC	1578
404	Qy	TyrAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGluGlnMetAlaSerAsp	423
1579	Db	TAT---GCCGGGAGGACCGAGGGAGGATGAGGAGCGCGAGACGGATGGCTTCTGAA	1635
424	Qy	ValAlaAsnAsnLysSerSerLeuGluAspGlyCysLeuSerCysGlyArgLysAsnPro	443
1636	Db	GTCACCAACAACAGGGCAATCTGGAAGACCGCTGTCTCTGTGGAAGAAAGAACCCCT	1695
444	Qy	ValSerPheHisProLeuPheGluGlyLysCysGlnThrCysArgAspArgPheLeu	463
1696	Db	GTGTCTTCCACCCCTCTTTGAGGGTGGGCTCTGTACAGTTTGCCGGGATCGCTTCCCTA	1755
464	Qy	GluLeuPheTyrMetTyrAspAspAspGlyTyrGlnSerTyrCysThrValCysCysGlu	483
1756	Db	GAGCTCTTTCATGTATGATGAGACCGCTATCATAGTCTCTATCTGCAACCGTGTCTGTGAG	1815
484	Qy	GlyArgGluLeuLeuLeuCysSerAsnThrSerCysCysArgCysPheCysValGluCys	503
1816	Db	GGCCGTGAAGTCTGCTGTGCAGTAACACAAAGCTGCTGCAGATGCTTCTGTGTGGAGTGT	1875
504	Qy	LeuGluValLeuValGlyThrGlyThrAlaAlaGluAlaLysLeuGlnGluProThrSer	523
1876	Db	CTGAGAGTGCTGGTGGGCCAGGCACAGCTGAGATGCCCAAGCTGCAGGAACCCCTGGAGC	1935
524	Qy	CysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgArgLysAspTyrAsn	543
1936	Db	TGCTATATGTGCCTCCCTCAGCGCTGCCATGGGGTCTCCGACGCAAGGAAGATTGGAAAC	1995
544	Qy	ValArgLeuGlnAlaPhePheThrSerAspThrGlyLeu---GluTyrGluAlaProLys	562
1996	Db	ATGGCGCTGCAAGACTCTTCTACTGATCTCTGACCTGGAGAAATTTGAGCCACCACAAG	2055
563	Qy	LeuTyrProAlaIleProAlaAlaArgArgProIleArgValLeuSerLeuPheAsp	582
2056	Db	TTGTACCCAGCAATTCTCGACCCAAAGAGGGGCCCATTAGAGTCTCTCTCTGTGTGAT	2115
583	Qy	GlyIleAlaThrGlyTyrLeuValLeuLysGluLeuGlyIleLysValGlyLysTyrVal	602
2116	Db	GCAATTGCACGGGGTACTTGTGTCTCAAGAGTGGGTATTAAAGTGGAAAAGTACATT	2175
603	Qy	AlaSerGluValCysGluGluSerIleAlaValAlaGlyThrValLysHisGluGlyAsnIle	622
2176	Db	GCCTCCGAAGTCTGTGCAAGTCCATCGCTGTGGAACTGTTAAGCATGAAGGCCAGATC	2235

Qy	623	LysTyrValAsnAspValArgAsnIleThrLysIysAsnIleGluGluTrpGlyProPhe	642
Db	2236	AAATATGTCATTCAGCTCCGGAATAATCACCAAGAAAATATTGAAGAGTGGGGCCGGTTC	2295
Qy	643	AspLeuValIleGlyGlySerProCysAsnAspLeuSerAsnValAsnProAlaArgLys	662
Db	2296	GACTTGGTGATTGGTGGAGCCCATCATGATCTCTTAACGTCATCTCTGCCCGCAA	2355
Qy	663	GlyLeuTyrGluGlyThrGlyArgLeuPhePheGluPheTyrHisLeuLeuAsnTyrSer	682
Db	2356	GGTTTATATGAGGGACAGGAAGGCTCTCTTCGAGTTTTTACCACTTCGCTGAATTATACC	2415
Qy	683	ArgProLysGluGlyAspAspArgProPhePheTyrMetPheGluAsnValValAlaMet	702
Db	2416	CGCCCCAAGGAGGGCCACACCGTCCATTCTCTGGATGTTTCGAGATGTTGTGCCATG	2475
Qy	703	LysValGlyAspPheArgAspIleSerArgPheLeuGluCysAsnProValMetIleAsp	722
Db	2476	AAAGTGAATGACCAAGAAACACATCTCAAGATTCCTGTCATGTAACCCAGTGAATCGAT	2535
Qy	723	AlaIleLysValSerAlaAlaHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMet	742
Db	2536	GCCATCAAGGTGCTGCTGCTCACAGGCCCGGTACTTCTGGGGTAACCTACCCGGAATG	2595
Qy	743	AsnArgProValIleAlaSerLysAsnAspLysLeuGluLeuGlnAspCysLeuGluTyr	762
Db	2596	AACAGGCCGTGATGGCTTCAAGAAATGATAAGCTCGAGCTGCAGGATGCTCGGAGTTC	2655
Qy	763	AsnArgIleAlaLysLeuLysLysValGlnThrIleThrThrLysSerAsnSerIleLys	782
Db	2656	AGTAGCAGCAAGAAAGTTAAAGAAAGTGCACACAATAACCCACCAAGTCGAATCCATCAGA	2715
Qy	783	GlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysGluAspValLeuTrpCys	802
Db	2716	CAGGCGCAAAACCCAGCTTTCCCTGTAGTCATGAATGGCAGGACGCTTTTGTGGTGC	2775
Qy	803	ThrGluLeuGluArgIlePheGlyPheProValHisTyrThrAspValSerAsnMetGly	822
Db	2776	ACTGAGCTCGAAAGGATCTTCGGCTTCCCTGCTCACACCGACGTCGTCCAAACATGGGC	2835
Qy	823	ArgGlyAlaArgGlnLysLeuLeuGlyArgCysSerTrpSerValProValIleArgHisLeu	842
Db	2836	CGCGGCGCCGTCAGAAAGCTCTGGCAGGTCTCTGGAGTGTACCGGTATCAGACACCTG	2895
Qy	843	PheAlaProLeuLysAspTyrPheAlaCysGlu	853
Db	2896	TTTGCCCTTCGAAGGACTACTTTGCTGTGAA	2928
RESULT 11			
AY078427		4006 bp	linear
LOCUS	AY078427	4006 bp	mus musculus DNA cytosine methyltransferase 3b6 (Dnmt3b6) mRNA, complete cds; alternatively spliced.
DEFINITION	AY078427		
ACCESSION	AY078427		
VERSION	AY078427.1	GI:21655120	
KEYWORDS			house mouse.
SOURCE			Mus musculus
ORGANISM			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE			1 (bases 1 to 4006)
AUTHORS			Chen, T., Ueda, Y. and Li, E.
TITLE			Dnmt3a2 encoded by transcripts from an intronic promoter of Dnmt3a is expressed abundantly in ES cells and germ cells
JOURNAL			Unpublished
REFERENCE			2 (bases 1 to 4006)
AUTHORS			Chen, T., Ueda, Y. and Li, E.
TITLE			Direct Submission
JOURNAL			Submitted (08-FEB-2002) Cardiovascular Research Center, Massachusetts General Hospital, 149 13th Street, Charlestown, MA 02129, USA
FEATURES			Location/Qualifiers



```
source
1. 4006
/organism="Mus musculus"
/strain="129S4/SvJae"
/db_xref="taxon:10090"
/chromosome="2"
/map="84.0 cM"
1. 4006
/gene="Dmnc3b6"
269. 2659
/gene="Dmnc3b6"
/function="de novo methylation"
/notes="contains PHD and PWP domains; alternatively
spliced"
/codon_start=1
/product="DNA cytosine methyltransferase 3b6"
/protein_id="A185482.1"
/db_xref="GI:2155121"
/translation="MKGDSRHLEERGAAGYEETIYVNGNFSDDSDTKAPSPVLE
AICTEPVCTPETHGRSSSRUSKREVSLINYTDMDTGDDDEVDGSDILMPK
LIREKDTRTRESPAVTRHNSGTSLEKORASPRIIRGGRHVOEYEPVETPR
SRRRASASASTPVSADVEMEEVTPKSPTSPVSDODDDEGMDTQVDESRD
GDSDEVODKEFGDLYMGKIKGSMMWPMVVSWMKATSKROAMPBMRWOWPREDGK
SEISADKLVALGELFSQHPNLTATPNKLVSTRKAMHTLEKARVACRTFSSPESLED
QKPLMEWAGGFRKPTGIEGLPNKKQPVNKSVRSDSNLEPRRENKSRRTTN
DSASESPPKRLKTNSTYKGRGEDESERMASVNNKMLEDRCLSCGKNPVS
FHLPEFGGLCOSGRDLEFMYDEDDGYSCVCCGRELILGNTSCCRCPVEG
LELVAGATREDAKLQEPWSCMYCLPQCHGVLRKRDMMNRLODFTTTDPDEEFP
PKLPAIPAKRRPRTVLSLFDGATGLVLEKELIKVEKYLASECAESIAVGVKH
EGQIKVNDVRIKTIKKNIEMWPFDLVIGSPNDLSVNPARKGIEGTRGLPEFY
HLNLTNRPKEDNRPFFMFEENVAMKVNKDISRLACNPMVMDAIXVSAARARY
FMGNLPGMNRIRPEPATYTDVSNMGRGAKOLGSRMSVPVIRHLPALPKDYACE"

BASE COUNT 1001 a 1017 c 1075 g 913 t
ORIGIN

Alignment Scores:
Pred. No.: 7.69e-173 Length: 4006
Score: 3415.50 Matches: 660
Percent Similarity: 81.52% Conservative: 50
Best Local Similarity: 75.77% Mismatches: 68
Query Match: 74.80% Indels: 93
DB: Gaps: 11

US-09-720-086-8 (1-853) x AY078427 (1-4006)

QY 1 MetLysGlyAspThrArgHisLeuAsnGlyGluGluAspAlaGlyArgGluAspSer 20
Db 269 ATGAAGGAGAGACAGACACATCTGAATGAAGAAGAGGTGCCAGCGGATATGAGAGTGC 328
QY 21 IleLeuValAsnGlyAlaCysSerAspGlnSerSerAsp-----SerPro 35
Db 329 ATTATCGTTAATGGGAACCTTCAGTACCACTCTTCAGACACGAAGATGCTCCCTCACCC 388
QY 36 ProIleLeuGluAlaIle-----ArgThrProGluIleArgGlyArgArg 50
Db 389 CCAAGTCTTGGAAGCAATCTGCACAGACAGTGTGCACACACAGACACGAGCCGACAG 448
QY 51 SerSerSerArgLeuSerIleArgGluValSerSerLeuLeuSerIleThrGlnAspLeu 70
Db 449 TCAAGCTCCCGGCTGTCTTAAGAGGAGGTCTCACGCTTCTGAATTACACGACGACATG 508
QY 71 ThrGlyAspGlyAspGly-----GluAspGlyAspGlySerAspThrProVal 86
Db 509 ACAGAGAGATGGAGACAGAGATGATGAATGATGATGATGATGATGATGATGATGATGAT 565
QY 87 MetProIleLeuPheArgGlu-----ThrArgThrArgSerGluSerProAlaVal 103
Db 566 ATGCCAAAGCTCACCGGTGAGACCAAGACACACAGACGCGCTCTGAAAGCCCGGCTGTC 625
QY 104 ArgThrArgAsnAsnAsnSerValSerSerArgGluArgHisArgProSerProArgSer 123
Db 626 CGAACCCCGACATAGCAATGAGACCTCAAGCTTGAGAGAGCAAGAGACCTCCCCAGATC 685
QY 124 ThrArgGlyArgGlnGlyArgAsnHisValAspGluSerProValGluPheProAlaThr 143

686 ACCGAGGTGGCAGGAGCCGCCACCATGTGCAGAGTACCTGTGGATTTCCGGCTACC 745
QY 144 ArgSerLeuArgArgArgAlaThrAlaSerAlaGlyThrProTrpProSerProSer 163
Db 746 AGGTCTGGAGACGTCGAGCATGTCTTCAGCAAGACAGCCATGTCTATCCCTGCCACG 805
QY 164 SerIleLeuThrIleAspLeuThrAspThrGluAspThrHisGlyThrProGlnSer 183
Db 806 -----GTCACTTCATGAGAGACG-----ACACCTAAGAGC 838
QY 184 SerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyMetGluSerPro 203
Db 839 GTCACTACCCCATCACTTGACTTGAAGATGAGATGAGAGGGGATGATGATACCA 898
QY 204 GlnValGluAlaAspSerGlyAspGlyAspSerSerGlyTyrGlnAspGlyArgGluPhe 223
Db 899 CAGGTGATGCAGAGACGACAGATGAGACACAGACAGATCATGATGATTAAGAGTTT 958
QY 224 GlyIleGlyAspLeuValTrpGlyLysIleLysGlyPheSerTrpTrpProAlaMetVal 243
Db 959 GGAATAGTGACCTCTGTGGGAAAGATCAAGGCTTCTCTGTGGCTGCCATGTGTG 1018
QY 244 ValSerTrpLysAlaThrSerIleArgGlnAlaMetSerGlyMetArgTrpValGlnTrp 263
Db 1019 GTGCTCGAAGAACCACTTCAGACGACAGCCATCCCGGAATGCCGTGGATCACTGG 1078
QY 264 PheGlyAspGlyLysPheSerGluValSerAlaAspLysLeuValAlaLeuGlyLeuPhe 283
Db 1079 TTTCGTGATGGCAAGTTTCTGATGATCTCTGTCACAAATGTTGGTCTTGGGGCTTTC 1138
QY 284 SerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerIleArgLysAlaMetTyr 303
Db 1139 AGCCAGCATTTAATCTGCTACTTCAATAGCTGTTCTTATAGAGGCAAGCATGTAC 1198
QY 304 HisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSerSerProGlyAsp 323
Db 1199 CACACTCGAGAGAACCCAGGTTGAGCTGACCAAGACCTTCTCAGCAGCTTCGAGAG 1258
QY 324 SerIleGluAspGlnLeuLysProMetLeuGluTyrAlaHisGlyLysPheLysProThr 343
Db 1259 TCACTGGAGAGACAGTGAAGCCATGCTGAGTGGGCCACCGTGGCTTCAAGCTTACT 1318
QY 344 GlyIleGluGlyLeuLysProAsnAsnThrGlnProValAlaAsnLysSerIleValArg 363
Db 1319 GGGATCGAGGGGCTCAAAACCAAGAAAGCAACAGTGTATTAATCGAAGTGGCGT 1378
QY 364 ArgAlaGlySerArgLysLeuGluSerArgLysTyrGluAsnLysThrArgArgArgThr 383
Db 1379 CGTTCAGACAGTGAAGCACTTAAGAACCCAGAGACCGGAGAACAAAGTCGAAGACGACA 1438
QY 384 AlaAspAspSerAlaThrSerAspTyrCysProAlaProLysArgLeuLysThrAsnCys 403
Db 1439 ACCAATGACTGTGCTCTTGAG---TCCCCCCCAACCAAGCCCTCAAGCAATATAGC 1495
QY 404 TyrAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGluGlnMetAlaSerAsp 423
Db 1496 TAT---GGCGGAGAGACGAGGAGGAGATGAGAGAGCCGAGAACGATGCTTGTGAA 1552
QY 424 ValAlaAsnAsnLysSerSerLeuGluAspGlyCysLeuSerCysGlyArgLysAsnPro 443
Db 1553 GTCAACCAACCAAGGCAATCTGGAAGACCGCTGTTGTCTGTGGAAGAAAGAACCCCT 1612
QY 444 ValSerPheHisProLeuPheGluGlyLysLeuCysGlnThrCysArgAspArgPheLeu 463
Db 1613 GTGTCTTCCACCCCTCTTGAAGGTGGGCTCTGTCAAGATTCCCGGATGCTTCTCTTA 1672
QY 464 GluLeuPheIleTyrMetIleArgAspAspGlyTyrGlnSerIleTyrCysThrValCysCysGlu 483
Db 1673 GAGCTCTTCTATCATGATGATGATGAGACGCGTATCACTGTCACACCGTGTGGCTGTGAG 1732
QY 484 GlyArgGluLeuLeuLeuCysSerAsnThrSerCysArgArgCysPheCysValGluCys 503
```



1733	Db	GGCCGTGAAC	TGCTGCTGTGTCAGTAA	CACAAGCTGCTGCAGATGCTTCTGTGTGGAGTGT	1792
504	QY	LeuGluValLeuValGlyThrGlyThrAlaAlaGluAlaValLeuGlnGluProTyrPse	523		
1793	Db	CTGGAGGTGCTGGTGGCGCAGGCACAGCTGAGAGTGCCAAAGCTGCAGGAACCCCTGGAGC	1852		
524	QY	CysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgArgLysAspTyrAsn	543		
1853	Db	TGCTATATGTGCTCCCTCAGCGCTGCCATGGGTCCTCCGACCGCAGAAAGATTGGAAAC	1912		
544	QY	ValArgLeuGlnAlaPhePheThrSerAspThrGlyLeu---GluTyrGluAlaProLys	562		
1913	Db	ATGGCCCTGCAAGACTTCTTCTACTACTGATCTCGACTCGAAGAAATTTGAGCCACCACAAG	1972		
563	QY	LeuTyrProAlaIleProAlaAlaArgArgArgProIleArgValLeuSerLeuPheAsp	582		
1973	Db	TTGTACCAGCAATTCCTGCACCCAAAGAGGCCCATTAGAGTCTCTGTCTCTGTATTGAT	2032		
583	QY	GlyIleAlaThrGlyTyrLeuValLeuLysGluLeuGlyIleLysValGlyLysTyrVal	602		
2033	Db	GGAAATTGCAACGGGGTACTTGTGTCTCAAGGAGTTGGGTATTAAAGTGGAAAAAGTACATT	2092		
603	QY	AlaSerGluValCysGluGluSerIleAlaValGlyThrValLysHisGluGlyAsnIle	622		
2093	Db	GCTTCCGAAGTCTGTGCAGAGTCCATCGCTGTGGNACTGTTAAGCATGAGGCCAGATC	2152		
623	QY	LysTyrValAsnAspValArgAsnIleThrLysLysAsnIleGluGluTrpGlyProPhe	642		
2153	Db	AAATATGTCAATGACGTCCGAAAAATCACCAAGAAAAATATTGAAGAGTGGGCGCCGTTTC	2212		
643	QY	AspLeuValIleGlyGlySerProCysAsnAspLeuSerAsnValAsnProAlaArgLys	662		
2213	Db	GACTTGGTGATTGGTGGAAAGCCATGCATGATCTCTCTAAAGTCAATCTCTGCCCGCAA	2272		
663	QY	GlyLeuTyrGluGlyThrGlyArgLeuPhePheGluPheTyrHisLeuLeuAsnTyrSer	682		
2273	Db	GGTTTATATGAGGGCACAGGAGGCTCTTCTTCGAGTTTTACACATGCTGTAATTATACC	2332		
683	QY	ArgProLysGluGlyAspAspArgProPhePheTyrMetPheGluAsnValValAlaMet	702		
2333	Db	CGCCCCAAGGAGGGCGCAACCGTCCATCTTCTGGATGTTTCGAGAAATGTTGTGGCCATG	2392		
703	QY	LysValGlyAspLysArgAspIleSerArgPheLeuGluCysAsnProValMetIleAsp	722		
2393	Db	AAAGTGAATGACAGAAGAGACATCTCAAGATTCTCTGGCATGTAAACCCAGTGTGATCGAT	2452		
723	QY	AlaIleLysValSerAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMet	742		
2453	Db	GCCATCAAGGTCTCTGCTGCTCACAGGGCCCGGTACTTCTGGGTAAACCTACCCGGAATG	2512		
743	QY	AsnArgProValIleAlaSerLysAsnAspLysLeuGluLeuGlnAspCysLeuGluTyr	762		
2513	Db	AAC-----	2515		
763	QY	AsnArgIleAlaLysLeuLysLysValGlnThrIleThrThrLysSerAsnSerIleLys	782		
2515	Db	-----	2515		
783	QY	GlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysGluAspValLeuTrpCys	802		
2515	Db	-----	2515		
803	QY	ThrGluLeuGluArgIlePheGlyPheProValHisTyrThrAspValSerAsnMetGly	822		
2516	Db	-----AGCATCTTCGGCTTCCTGCTCACTACACGGACGTGTCCAACATGGGC	2563		
823	QY	ArgGlyAlaArgGlnLysLeuLeuGlyArgSerTrpSerValProValIleArgHisLeu	842		
2564	Db	CGCGCGCCCGTCAGAAAGCTGCTGGCGAGTCTCTGAGTGTACCGGTCTATCAGACACCTG	2623		
843	QY	PheAlaProLeuLysAspTyrPheAlaCysGlu	853		
2624	Db	TTTGCCCGCTTGAGGACTACTTTCCCTGTGAA	2656		



QY 723 AlalileValserAlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMet 742  
Db 2481 GCCATCAGGTGCTGCTCTCACAGGCCCGGTACTTCTGGGGTAACTTACCCGGGATG 2540  
QY 743 AsnArgProValileAlaSerLysAsnAspLysLeuGluLeuGlnApsCysLeuGluTyr 762  
Db 2541 AAC----- 2543  
QY 763 AsnArgIleAlaLysLeuLysLysValGlnThrIleThrThrLysSerAsnSerIleLys 782  
Db 2543 ----- 2543  
QY 783 GlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysGluAspValLeuTrpCys 802  
Db 2543 ----- 2543  
QY 803 ThrGluLeuGluArgIlePheGlyPheProValHisTyrThrAspValSerAsnMetGly 822  
Db 2544 -----AGGATCTTGGGCTTCCCTGCTCACTACACGGGAGCTGTCCAACATGGGC 2591  
QY 823 ArgGlyAlaArgGlnLysLeuLeuGlyArgSerTrpSerValProValIleArgHisLeu 842  
Db 2592 CCGCGCGCCGTCAGAGAGCTGTGGGAGGCTCTGGAGTGTACCGGTATCAGACACCTG 2651  
QY 843 PheAlaProLeuLysAspTyrPheAlaCysGlu 853  
Db 2652 TTTGCCCCCTTGAAGGACTACTTTGCTGTGAA 2684  
RESULT 13  
AF151972 4149 bp mRNA linear ROD 08-JUN-2000  
LOCUS AF151972 Mus musculus DNA cytosine-specific methyltransferase isoform 4  
DEFINITION AF151972.1 GI:8347125  
ACCESSION AF151972  
VERSION AF151972.1  
KEYWORDS Mus musculus  
SOURCE Mus musculus  
ORGANISM Mus musculus  
REFERENCE 1 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
Yin, B., Chen, Y.T., Zhu, M., Luo, Y.J., Zhu, N., Xu, S.C., Wu, G.Y. and  
Shen, Y.  
TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing  
isoforms in mouse embryonic tissue  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 4149)  
AUTHORS Yin, B., Chen, Y.T., Zhu, M., Luo, Y.J., Zhu, N., Xu, S.C., Wu, G.Y. and  
Shen, Y.  
TITLE Direct Submission  
JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular  
Biology, Institution of Basic Medical Sciences, Chinese Academy of  
Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,  
P.R.China  
FEATURES  
source  
1..4149  
/organism="Mus musculus"  
/strain="KM"  
/db\_xref="taxon:10090"  
/dev\_stage="8-9 day old embryo"  
1..4149  
/gene="Dnmt3b"  
412..2802  
/gene="Dnmt3b"  
/function="methylates cytosine in DNA"  
/note="alternatively spliced"  
/codon\_start=1  
/product="DNA cytosine-specific methyltransferase isoform  
4"  
/protein\_id="AAF74518.1"  
/db\_xref="GI:8347126"  
/translation="MKGSRHLNBEAGSGYBECLIVNGNFSDQSDTKDAPPPVL  
AICTEPVCTPETRRSSRLSKREVSSLLNTYQDMIGDRODDEVDGNGSDILMPK

BASE COUNT 1013 a 1082 c 1123 g 931 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3,08e-172 Length: 4149  
Score: 3404.50 Matches: 659  
Percent Similarity: 81.40% Conservative: 50  
Best Local Similarity: 75.66% Mismatches: 69  
Query Match: 74.56% Indels: 93  
DB: 10 Gaps: 11  
US-09-720-086-8 (1-853) x AF151972 (1-4149)  
QY 1 MetLysGlyAspThrArgHisLeuAsnGlyGluGluAspAlaGlyGlyArgGluAspSer 20  
Db 412 ATCAAGGGAGACAGACACATCTGAATGAAGAGAGGGTCCAGCGGTATGAGGAGTGC 471  
QY 21 IleLeuValAsnGlyAlaCysSerAspGlnSerSerAsp-----SerPro 35  
Db 472 ATATCTGTTAATGGAACCTTCAGTGACCATCTCTCAGACACAGAGGATGCTCCCTCACC 531  
QY 36 ProIleLeuGluAlaIle-----ArgThrProGluIleArgGlyArgArg 50  
Db 532 CCAGTCTTGGAGGCAATCTCCACAGAGCCAGTCTGCACACAGAGAGACAGAGCGCGCAG 591  
QY 51 SerSerSerArgLeuSerLysArgGluValSerSerLeuLeuSerTyrThrGlnAspLeu 70  
Db 592 TCAAGTCTCCGGCTGCTTAAGAGGGAGTCTCCAGCCTCTTCTGAATTACCGCAGGACATG 651  
QY 71 ThrGlyAspGlyAspGly-----GluAspGlyAspGlySerAspThrProVal 86  
Db 652 ACAGAGAGTGGAGACAGAGATGATGAAGTAGATGATGGAGTATGCTCTGATATT---CTA 708  
QY 87 MetProLysLeuPheArgGlu-----ThrArgThrArgSerGluSerProAlaVal 103  
Db 709 ATGCCAAAGTCAACCGTGTAGACCAAGGACACAGGACCGCTCTGAAAGCGCGCTGTC 768  
QY 104 ArgThrArgAsnAsnAsnSerValSerSerArgGluArgHisArgProSerProArgSer 123  
Db 769 CGAACCCTGATAGCATGGACCTCCAGCTTGGAGAGGCAAGAGCCTCCCCAGAAATC 828  
QY 124 ThrArgGlyArgGlnGlyArgAsnHisValAspGluSerProValGluPheProAlaThr 143  
Db 829 ACCCGAGGTCCGAGGCGCGCCACCATGTGCAGGAGTACCTGTGGAGTTCCTCGGCTACC 888  
QY 144 ArgSerLeuArgArgArgAlaThrAlaSerAlaGlyThrProTyrProSerProProSer 163  
Db 889 AGGTCTCGGAGAGCTCGAGCATCGTCTTCAAGACAGCAGCATGTCATCCCTCCACG 948  
QY 164 SerTyrLeuThrIleAspLeuThrAspAspThrGluAspThrHisGlyThrProGlnSer 183  
Db 949 -----GTCGACTTTCATGGAAGAGT-----ACACCTAAGAGC 981  
QY 184 SerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyGlyMetGluSerPro 203  
Db 982 GTCAGTACCCCATCACTTGACTTGAAGCAGGATGGAGATCAGGAGGATATGATACACCA 1041  
QY 204 GlnValGluAlaAspSerGlyAspGlyAspSerSerGluTyrGlnAspGlyLysGluPhe 223  
Db 1042 CAGGTGGATGACAGAGACATATATGGAGACAGCAGCATATCAGGATATGATTAAGAGT 1101

QY 224 G1Y1IeGIAspLeuValTPG1YLYS1IeLYSG1YpHeSerTPTrProAlaMetVal 243  
 Db 1102 GGAATTAGGTGACCTCGTGTGGGAAAGATCAAGGGCTTCTCTGCTGGCTGCCATGTG 1161  
 QY 244 ValSerTPLeValaThrSerLYsArgGlnAlaMetSerGlyMetArgTPValGlnTP 263  
 Db 1162 GTGCTCTGGAAAGCCACTTCCAGCGACAGCGCCATGCCGAATCGCTGGGACACTGG 1221  
 QY 264 PheG1YAspG1YLYsPheSerG1YValSerAlaAspLYsLeuValAlaLeuG1YLeuPhe 283  
 Db 1222 TTGGTGTGAGGAAGTTTCTTGAGATCTCTGCTGACAACTGTGTGCTTGGGGCTTTC 1281  
 QY 284 SerGlnHisPheAsnLeuAlaThrPheAsnLYsLeuValSerTYrArgLYsAlaMetTYr 303  
 Db 1282 AGCCAGACCTTAAATCTGCTACCTTCAATAGCTGTTCTTATAGGAAGGCCATGTAC 1341  
 QY 304 HisAlaLeuG1YLYsAlaArgValaArgAlaG1YLYsThrPheProSerSerProG1YAsp 323  
 Db 342 CACACTCTGGAGAAAGCCAGGGTTCAGCTGGCAGACCTTCTCCAGCGCTCTGGAGAG 1401  
 QY 324 SerLeuG1YAspG1YLeuLYsProMetLeuG1YTPAlaHisG1YLYsPheLYsProThr 343  
 Db 1402 TCACCTGGAGGACACGCTGAAGCCATGCTGGAGTGGGCCCGCTTCAAGCTTACT 1461  
 QY 344 G1Y1IeGIYLYsLYsProAsnAsnThrGlnProValValaAsnLYsSerLYsValArg 363  
 Db 1462 GGGATCGAGGGCTCAAAACCCACAGAAAGCAACAGTGTATTAAGTCGAAGGTGCT 1521  
 QY 364 ArgAlaG1YSerArgLYsLeuG1YLYsArgLYsTYrG1YLYsLeuAsnLYsThrArgArgThr 383  
 Db 1522 CGTTCAGACAGTGAAGAACTTGAACCCAGAGACGCGAAGAAACAAAGTGAAGACGACA 1581  
 QY 384 AlaAspAspSerAlaThrSerAspTYrCysProAlaProLYsArgLeuLYsThrAsnCys 403  
 Db 1582 ACCAATGACTGCTGCTTCTGAG---TCCCCCCCCCAAGCGCTCTCAAGACAAATAGC 1638  
 QY 404 TYrAsnAsnG1YLYsAspArgG1YAspG1YAspG1YAspArgG1YMetAlaSerAsp 423  
 Db 1639 TAT---GGCGGGAAGACCGAGGGAGATGAAGAGACCCAGAACCGATGCTTCTGAA 1695  
 QY 424 ValAlaAsnAsnLYsSerSerLeuG1YAspG1YLYsLeuSerCysG1YArgLYsAsnPro 443  
 Db 1696 GTCCACCAACACAGAGGCAATCTGAAGACCGCTTGTCTCTGCGAAGAAAGAACCTT 1755  
 QY 444 ValSerPheHisProLeuPheG1YLYsG1YLYsCysGlnThrCysArgAspArgPheLeu 463  
 Db 756 GTGCTCTTCCACCCCTCTTGGAGGTGGCTCTGTCAAGTTCGGGATCCCTTCTTA 1815  
 QY 464 GluLeuPheTYrMetTYrAspAspArgG1YTYrGlnSerTYrCysThrValCysCysGlu 483  
 Db 1816 GAGCTCTTCTACATGATGATGAGAGCGGCTATCAGTCTTACTGCAACCGTGTGTGAG 1875  
 QY 484 G1YArgG1YLeuLeuLeuCysSerAsnThrSerCysCysArgCysPheCysValAlaCys 503  
 Db 1876 GGCGGTGAAGCTGCTGCTGTGAGTAACACAGCTGCTGCAGATGCTTCTGTGTGAGGT 1935  
 QY 504 LeuGluValAlaLeuValG1YThrG1YThrAlaAlaG1YAlaLYsLeuGlnGluProTPSer 523  
 Db 1936 CTGAGAGTGTGTGGCGGAGGACAGCTGAGAGATGCCAAGCTGCAGGAACCTTGAGGC 1995  
 QY 524 CysTYrMetCysLeuProGlnArgCysHisG1YValaLeuArgArgLYsAspTPAsn 543  
 Db 1996 TGCATATATGTCTCTCCCTCAAGCGTGCATGGGTCTCTCCGACGAGAAAGATTGGAAC 2055  
 QY 544 ValArgLeuGlnAlaPhePheThrSerAspThrG1YLeu---GluTYrG1YValaProLYs 562  
 Db 2056 ATGGCCCTGCAAGACTTCTTCACTATGATCTGACCTGGAAGAAATTGGCCACCCCAAG 2115  
 QY 563 LeuTYrProAlaAlaPheProAlaAlaArgArgArgProAlaArgValaLeuSerLeuPheAsp 582  
 Db 2116 TTGTATCCCGACAACTTCTGCGAGCCAAAGAGAGGCCCATTTAGATGCTGTCTGTTCAT 2175  
 QY 583 GlyIleAlaThrG1YTYrLeuValLeuLYsG1YLeuG1YIleLYsValG1YLYsTYrVal 602

Db 2176 GGAATTGACAGCGGGTACTTGTGCTCAAGAGTGGTGTAAAGTGAAGAAAGTACATT 2235  
 QY 603 Alaserg1YValCysGluGluSerIleAlaValG1YThrValLYsHisGluG1YAsnIle 622  
 Db 2236 GCCCTCCAGAGTGTGTGACAGTCCATCGCTGTGGGAAGCTTTAAGCAAGAAAGCCAGATC 2295  
 QY 623 LYsTYrValAsnAspValArgAsnIleThrLYsAsnIleGluG1YTPG1YProPhe 642  
 Db 2296 AAATATGTCATGAGCTCCGGAATAATCCAAAGAAATAATTGAAGATGGGGCCGCTTC 2355  
 QY 643 AspLeuValIleG1YLYsSerProCysAsnAspLeuSerAsnValAsnProAlaArgLYs 662  
 Db 2356 GACTTGTGATGTGTGGAGCCCATGCAATGATCTCTTAACGTCATCTGCCCGGAAA 2415  
 QY 663 G1YLeuTYrG1YLYsThrG1YArgLeuPhePheGluPheTYrHisLeuLeuAsnTYrSer 682  
 Db 2416 GGTATATAGAGGCGACAGGAAGCTCTTCTTGAATTTACCACTTGTGAATTTATACC 2475  
 QY 683 ArgProLYsG1YLYsAspAspArgProPhePheTYrMetPheGluAsnValAlaMet 702  
 Db 2476 CGCCCAAGAGGGGACACACCGTCCATCTTCTGATGTTCCAGAAATGTTGGCCATG 2535  
 QY 703 LYsValG1YAspLYsArgAspIleSerArgPheLeuGluCysAsnProValMetIleAsp 722  
 Db 2536 AAAGTGAATGACAGAAAGATCTCAAGATCTCTGCGATGTAACCCAGTATGATGAT 2595  
 QY 723 AlaIleLYsValSerAlaAlaHisArgAlaArgTYrPheTPTrG1YAsnLeuProG1YMet 742  
 Db 2596 GCCATCAAGGTGTCTGCTGCTCAAGAGCCCGGTACTTCTGGGGTAACTTACCCGGAATG 2655  
 QY 743 AsnArgProValIleAlaSerLYsAsnAspLYsLeuG1YLeuGlnAspCysLeuGluTYr 762  
 Db 2656 AAC----- 2658  
 QY 763 AsnArgIleAlaLYsLeuLYsValGlnThrIleThrLYsSerAsnSerIleLYs 782  
 Db 2658 ----- 2658  
 QY 783 GlnG1YLYsAsnG1YLeuPheProValValMetAsnG1YLYsGluAspValLeuTPCys 802  
 Db 2658 ----- 2658  
 QY 803 ThrGluLeuGluArgIlePheG1YPheProValHisTYrThrAspValSerAsnMetGly 822  
 Db 2659 -----AGATCTTCCGGCTTCCCTGCTCACTACACGAGAGTGCACCAATGAGGC 2706  
 QY 823 ArgG1YAlaArgG1YLYsLeuLeuG1YLYsSerTPSerValProValIleArgHisLeu 842  
 Db 2707 CGCGGCGCCGTCAGAAAGTGTGGGAGGTCTGTGAGTGAAGTGAACGGTCAAGACACCTG 2766  
 QY 843 PheAlaProLeuLYsAspTYrPheAlaCysGlu 853  
 Db 2767 TTTCCTCCCTTGAAGACTACTTGTGCTGTGAA 2799  
 RESULT 14  
 AF068628 3946 bp mRNA linear ROD 06-DEC-1999  
 AF068628  
 LOCUS Mus musculus DNA cytosine-5 methyltransferase 3B3 (Dnm3b) mRNA,  
 DEFINITION alternatively spliced, complete cds.  
 ACCESSION AF068628.2 GI:6449473  
 VERSION  
 KEYWORDS  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 3946)  
 AUTHORS Okano,M., Xie,S. and Li,E.  
 TITLE Cloning and characterization of a family of novel mammalian DNA  
 (cytosine-5) methyltransferases  
 JOURNAL Nat. Genet. 19 (3), 219-220 (1998)  
 MEDLINE 98324766

PUBMED 9662389  
REFERENCE 2 (bases 1 to 3946)  
Xie, S., Okano, M. and Li, E.  
TITLE Direct Submision  
JOURNAL Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,  
Charlestown, MA 02129, USA  
REFERENCE 3 (bases 1 to 3946)  
AUTHORS Okano, M., Chijiwa, T., Sasaki, H. and Li, E.  
TITLE Direct Submision  
JOURNAL Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,  
Charlestown, MA 02129, USA  
REMARK Sequence update by submitter  
COMMENT On Nov 18, 1999 this sequence version replaced gi:3327983.  
FEATURES  
source  
1..3946  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/map="84.0 cM"  
1..3946  
/gene="Dnmt3b"  
misc\_feature  
1..3946  
/gene="Dnmt3b"  
/note="similar to EST sequences deposited in GenBank  
Accession Numbers AA116694, AA119979, AA177277, AA210568,  
AA407106, and AA575617"  
269..2699  
/gene="Dnmt3b"  
/function="de novo DNA methylation"  
/note="alternatively spliced product; contains Cys-rich  
region"  
/codon\_start=1  
/product="DNA cytosine-5 methyltransferase 3B3"  
/protein\_id="AAC40180.2"  
/db\_xref="GI:6449474"  
/translation="MKGDSRLNREAGSGVECLIVNGNFSDDSTKDAPSPVLE  
AICTEPCTPEIRGRSSRLSKREVSSLLNTQDMTGDDRDDEVDGNGSILMPK  
LRRKQDTRRSESPAVRHSNGTSSLRQSRASPRITGRQRHVVQETPEVFPATR  
SRRRSASSTPWSPASVDFMEVTPKSVSPVSDLSQDQDQGMDFVQDAESRD  
GDSTYQDDKEFGIDLVGKIKGFSWPMVMSWKATSKROMPMGRWVQWFGDKF  
SEISADKLVALGFLSQHFNLFATFNKLSYRKAMHTLEKARVAGRTFSSPQESLED  
OLKMLEWAHGGFKPTGIEGLKPNKKQPKNSRRRTNDSSAESPCKRLKTNISYGG  
KDRDESRESRMASVETNGNKLNDRLCSCKNPNVSPHPLPEGLCQSCROFLEL  
FYMYDEGYOSYCTVCCSEGLLLCSNCCRCFCVECLVLVGLGCTADAKLQEPWS  
CYMCLPQCHVLRKKDNNRLOPFTDPLEPEPKLYPAIPAAKRPRLVLSL  
FDGATGLVLUKEIGIKVEKTIASVCAESIAVGVKHEGQIKYNDVRKITKNIEE  
WGPDVLVIGSPCNDLSNVNPAKGLYEGTGRLEFFEFYHLLNTRPREGDNRRPFWMF  
ENVVAMKYNDDKDISRFLACNPVMTDAIKVSAARARYFWGNLPGMNRIFGFPFAHYTD  
VSNMGRGARQKLLGRSWSVPVIRHLFAPLKDYFACE"  
BASE COUNT 983 a 1006 c 1056 g 901 t  
ORIGIN

Alignment Scores:  
Pred. No.: 2,22e-168 Length: 3946  
Score: 3331.50 Matches: 646  
Percent Similarity: 79.68% Conservative: 48  
Best Local Similarity: 74.17% Mismatches: 64  
Query Match: 72.96% Indels: 113  
DB: 10 Gaps: 12

US-09-720-086-8 (1-853) x AF068628 (1-3946)

QY 1 MetLysGlyAspThrArgHisLeuAsnGlyGluAspAlaGlyArgGluAspSer 20  
Db 269 ATGAAGGAGACAGACAGACATCTGAATGAAGAAGAGCGGTGCCAGCGGTATGAGAGTGC 328  
QY 21 IleLeuValAenGlyAlaCysSerAspGlnSerSerAsp-----SerPro 35  
Db 329 ATTATCGTTAATGGCAACTTCAGTGACCATCTCTCAGACACGAAAGGATGCTCCCTCACCC 388  
QY 36 ProIleLeuGluAlaIle-----ArgThrProGluIleArgGlyArg 50  
Db 389 CCAGTCTTGGAGGCANTCTGCAACAGAGCCAGTCTGCACACCAAGAGCCAGAGCCGCGCAGG 448

QY 51 SerSerSerArgLeuSerLysArgGluValSerSerLeuSerTyrThrGlnAspLeu 70  
Db 449 TCAAGCTCCCGGCTGCTAAGAGGGAGGCTCTCAGACCTTCTGAATATCACCGAGACATG 508  
QY 71 ThrGlyAspGlyAspGly-----GluAspGlyAspGlySerAspThrProVal 86  
Db 509 ACAGGAGATGGAGACAGACATGATGAAGTAGATGATGGAATGCTCTGATATT---CTA 565  
QY 87 MetProLysLeuPheArgGlu-----ThrArgThrArgSerGluSerProAlaVal 103  
Db 566 ATGCCAAAGCTCACCGTGAACCAAGGACACAGGAGCGCTCTGAAAGCCGCGCTGTC 625  
QY 104 ArgThrArgAsnAsnAsnSerValSerSerArgGluArgHisArgProSerProArgSer 123  
Db 626 CGAACCCGACATAGCAATGGGACCTCCAGCTTGAGAGGCAAGAGCGCTCCCCCAAGATC 685  
QY 124 ThrArgGlyArgGlnGlyArgHisValAspGluSerProValGluPheProAlaThr 143  
Db 686 ACCCGAGGTGGGAGGCCGCCACCATGTGCAGAGTAGTACCTGTGGAGTTTCCGCTACC 745  
QY 144 ArgSerLeuArgArgAlaThrAlaSerAlaGlyThrProThrProSerProSerPro 163  
Db 746 AGGTCTCGGAGACGTCGACATCGTCTTCAGCAAGCACGCGCATGCTCATCCCTGCCAGC 805  
QY 164 SerTyrLeuThrIleAspLeuThrAspAspThrGluAspThrHisGlyThrProGlnSer 183  
Db 806 -----GTCGACTTCATGGAAGAGTG-----ACACCTAAGAGC 838  
QY 184 SerSerThrProTyrAlaArgLeuAlaGlnAspSerGlnGlnGlyMetGluSerPro 203  
Db 839 GTCACTACCCCATCATGTTGACTTGAGCCAGGATCGAGATCAGGAGGTTATGATACCA 898  
QY 204 GlnValGluAlaAspSerGlyAspGlyAspSerSerGluTyrGlnAspGlyLysGluPhe 223  
Db 899 CAGGTGGATGCAGAGACAGACAGATGGAGACAGACAGATATCAGAGTATGATAAGAGTTT 958  
QY 224 GlyIleGlyAspLeuValThrGlyLysIleLysGlyPheSerThrProAlaMetVal 243  
Db 959 GGAATAGGTGACCTCGTGTGGGAAAGATCAAGGGCTTCTCCTGCTGGCTGCCATGTG 1018  
QY 244 ValSerTrpLysAlaThrSerLysArgGlnAlaMetSerGlyMetArgTrpValGlnTrp 263  
Db 1019 GTGCTCTGGAAGCCACCTCCAGACGACAGCCATGCCGGAATGCCGTGGTACAGTGG 1078  
QY 264 PheGlyAspGlyLysPheSerGluValSerAlaAspLysLeuValAlaLeuGlyLeuPhe 283  
Db 1079 TTTGGTGTATGGCAAGTTTCTGAGATCTCTGCTGACAAACTGGTGGCTCTGGGGCTGTC 1138  
QY 284 SerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyrArgLysAlaMetTyr 303  
Db 1139 AGCCAGCACTTTAATCTGGCTACCTTCAATAAGCTGTTCTTATAGGAAGCCCATGTAC 1198  
QY 304 HisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSerSerProGlyAsp 323  
Db 1199 CACACTCTGGAGAAAGCCAGGTTTCGAGCTGGCAAGACCTTCTCCAGCAGTCCCTGGAGAG 1258  
QY 324 SerLeuGluAspGlnLeuLysProMetLeuGluTrpAlaHisGlyGlyPheLysProThr 343  
Db 1259 TCACCTGGAGGACAGCTGAAGCCCATGCTGGAGTGGGCCACAGCTGGCTTCAAGCTACT 1318  
QY 344 GlyIleGluGlyLeuLysProAsnAsnThrGlnProValValAsnLysSerLysValArg 363  
Db 1319 GGGATCGAGGGCTCAAAACCAACCAAGCAACCA----- 1354  
QY 364 ArgAlaGlySerArgLysLeuGluSerArgLysTyrGluAsnLysThrArgArgThr 383  
Db 1355 -----GAGAACAAAAGTCTGAAGACGCACA 1378  
QY 384 AlaAspAspSerAlaThrSerAspTyrCysProAlaProLysArgLeuLysThrAsnCys 403  
Db 1379 ACCAATGACTCTGCTCTCTGAG---TCCCCCCCCACCAAGCGCCTCAAGACAAATAGC 1435

QY 404 TyrAsnaNGlLYsAspaRgGlYAspGlInSerArgGlInMeAlaSerAsp 423  
 Db 1436 TAT--GGCGGGAAGAGCCGAGGAGATGAGAGAGCCAGAAACGATGCTTCGAA 1492  
 QY 424 ValAlaAsnaSrlSerSerLeuGlInuAspGlyCysLeuSerCysGlyArgLYsAsnPro 443  
 Db 1493 GTCCACCAACAAAGGGCAATCTGAGAGACCGCTGTTGCTCTGGGAAAGAAACCTT 1552  
 QY 444 ValSerPheHisProLeuPheGluGlyLeuCysGlnThrCysArgAspArgPheLeu 463  
 Db 1553 GTGCTCTTCACCCCTCTTTAGAGGTGGCTCTGTCAGAGTTGCGGATCCCTTCTTA 1612  
 QY 464 GluLeuPheTyMetTyrAspAspAspGlyTyrGlnSerTyrCysThrValCysCysGlu 483  
 Db 1613 GAGCTCTTCAATGATGATGAGAGACCGCTATCACTCACTGACCGCTGCTGTAG 1672  
 QY 484 GlYArgGlulLeuLeuLeuCysSerAsnThrSerCysCysArgCysPheCysValGluCys 503  
 Db 1673 GGGCGTGAACCTGCTGCTGCGAGTACACAAAGCTGCGAGATGCTTCTGTGTGAGTGT 1732  
 QY 504 LeuGluValLeuValGlyThrGlyThrAlaAlaGluAlaLysLeuGlnGluProTyrSer 523  
 Db 1733 CTGAGAGTCTGTGTGGCGGACGACAGCTGAGAGATCCAGCTGAGAAATTTAGCCACCC 1792  
 QY 524 CysTyrMetCysLeuProGlnArgCysHisGlyValLeuArgArgArgLYsAspTyrAsn 543  
 Db 1793 TGCCTATATGCTCTCCCTCAGCGCTGCGCATGGGGTCTCCAGCGAGGAAAGATTGGAAC 1852  
 QY 544 ValArgLeuGluAlaPhePheThrSerAspThrGlyLeu--GluTyrGluAlaProLys 562  
 Db 1853 ATGGCGCTGCAAGACTTCTTCACTACTGATCTCGACCTGGAAGATTTAGCCACCCAG 1912  
 QY 563 LeuTyrProAlaIleProAlaAlaArgArgArgProIleArgValLeuSerLeuPheAsp 582  
 Db 1913 TTGTACCCACGCAATTCCTGCGAGCCAAAGAGGCCCATTTAGAGTCTGCTCTGTTGAT 1972  
 QY 583 GlyTLeaIaThrGlyTyrLeuValLeuLysGluLeuGlyIleLysValGlyLysTyrVal 602  
 Db 1973 GGAATTGCAACGGGATCTGTGCTGCTCAAGAGAGTTGGGTATTAAAGTGAAGAAATGACT 2032  
 QY 603 AlaSerGluValCysGluGlnSerIleAlaValGlyThrValLysHisGluGlyAsnIle 622  
 Db 2033 GCCTCCGAAGTCTGTGAGAGTCCATCGCTGTGGAACTGTTAAGCATTAAGGCCACAGTC 2092  
 QY 623 LysTyrValAsnAspValArgAsnIleThrLysLysAsnIleGluGluTyrGlyProPhe 642  
 Db 2093 AAATATGTCATGACGCTCCGGAATAATCACAGAAATAATTGAAGAGTGGGCCCTTC 2152  
 QY 643 AspleuValIleGlyGlySerProCysAsnAspLeuSerAsnValAsnProAlaArgLys 662  
 Db 2153 GACTTGGTATGCTGTGAGAGCCCATGATGATCTCTTAACGTCAATCTGCGCCGCAAA 2212  
 QY 663 GlyLeuTyrGluGlyThrGlyArgLeuPhePheGluPheTyrHisLeuLeuAsnTyrSer 682  
 Db 2213 GCTTATATGAGGCGACAGAGAGCTCTTCTTGAAGTTTACCACTTGTCTGAATTATACC 2272  
 QY 683 ArgProLYsGluGlyAspAspArgProPhePheTyrMetPheGluAsnValValAlaMet 702  
 Db 2273 CGCCCAAGGAGGCGACACGCTCATCTTCTGATGTTGAGAGATGTTGTGGCCATG 2332  
 QY 703 LysValGlyAspLYsArgAspIleSerArgPheLeuGluCysAsnProValMetIleAsp 722  
 Db 2333 AAAAGTGAATGACAAAGAAACATCTCAAGATTCCTGGCATGTAAACCAAGATGATGAT 2392  
 QY 723 AlaIleLysValSerAlaIleHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMet 742  
 Db 2393 GCCATCAAGGTGTCTGTGTCAACAGGCCCGGTACTTCTGGGGTAACTTAACCCGGAATG 2452  
 QY 743 AsnArgProValIleAlaSerLysAsnAspLysLeuGluAspCysLeuGluTyr 762  
 Db 2453 AAC----- 2455  
 QY 763 AsnArgIleAlaLysLeuLysValGlnThrIleThrThrLysSerAsnSerIleLys 782

Db 2455 ----- 2455  
 QY 783 GlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLYsGluAspValLeuTyrCys 802  
 Db 2455 ----- 2455  
 QY 803 ThrGluLeuGluArgIlePheGlyPheProValHisTyrThrAspValSerAsnMetGly 822  
 Db 2456 -----AGGATCTTCGGCTTCCCTGCTCACTACAGAGAGTGTCAACATGGGC 2503  
 QY 823 ArgGlyAlaArgGlnLysLeuLeuGluYArgSerTyrSerValProValIleArgHisLeu 842  
 Db 2504 CGCGCGCCCTCAGAGCTGCTGGGACAGTCTCTGAGATGTACCGGTATCAGACACCTG 2563  
 QY 843 PheAlaProLeuLysAspTyrPheAlaCysGlu 853  
 Db 2564 TTTGCCCTTCAAGACTACTTTCCTGTGAA 2596  
 RESULT 15  
 AF151976 3974 bp mRNA linear ROD 08-JUN-2000  
 LOCUS AF151976 Mus musculus DNA cytosine-specific methyltransferase isoform 8  
 DEFINITION (Dmmt3b) mRNA, complete cds.  
 ACCESSION AF151976  
 VERSION AF151976.1 GI:8347136  
 KEYWORDS Mus musculus.  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 REFERENCE 1 (bases 1 to 3974)  
 AUTHORS Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and Shen, Y.  
 TITLE Cloning of full-length Dmmt3b cDNA and its alternative splicing isoforms in mouse embryonic tissue  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 3974)  
 AUTHORS Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and Shen, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular Biology, Institution of Basic Medical Sciences, Chinese Academy of Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005, P. R. China  
 FEATURES  
 source  
 location/Qualifiers  
 1..3974  
 /organism="Mus musculus"  
 /strain="XM"  
 /db\_xref="taxon:10090"  
 /dev\_stage="8-9 day old embryo"  
 1..3974  
 /gene="Dmmt3b"  
 297..2627  
 /gene="Dmmt3b"  
 /function="methylates cytosine in DNA"  
 /note="alternatively spliced"  
 /codon\_start=1  
 /product="DNA cytosine-specific methyltransferase isoform 8"  
 /protein\_id="AAF74522.1"  
 /db\_xref="GI:8347137"  
 /translation="MKGDSRLNEEGASGYEECIIVNGNSDOSDSTKQAPPVLE  
 AICTEPVCTPRTGRSSSRSLKREYSSLLNTYODMKGDBDDEDDGSGDLMPK  
 LTRTKDTRTRSSPAVTRHNSGTSLSLRQASPRITRGGRHHVQEVPERPATR  
 SRRRASSSASTPWSVDPMEEVTPKSVSTPSVDLSODGDEMDTTOVDAESLY  
 GDSSTYODDKKEFGIDLVWGIKGFSPWAMVYSWKATSKQAMPGRWQWFGDGK  
 SEISADKVALGIFSOHFNLTATFKLVSRKAMHTLEKARYRAGCTPSSPGSLED  
 QLKPELWHAHGFKPTGIEGLKPNRKQEPNKSRRRTNDSASSEPPPKLKTNSYGG  
 KDRGEDEERERMASEVTNNKGNLEDRCLSCGKRPVSFHPLEBGLCOGRFLFL  
 FYMVDGYSYCTVCEGRELILCSNCCRCFCVGLVVGAGTAEDAKLOEPMW  
 CYMCLPQRCHGVARRKDMNRLODPTTDPDLSEFPKLYPAIPAAKRRPIRLVSL  
 FDGIATGYLVKELGIVKRYIASEVCASIAVGTVAHEGQIKYVNDVRKITKNIEE



WGPFDLVIGGSPCNLSNVNPKARKGLYEGTGRLEFFBYHLLNLTYPKKGDNRRPFWMF  
ENVNMYKNDKDIISFLACNPVMDIAIKVSAHRARYFWGNLPNMRIFGFPAYTD  
VSNMGRGARQKLLGRSWSVFIHLFAPLKYDFACE"  
BASE COUNT 972 a 1032 c 1078 g 892 t  
ORIGIN

Alignment Scores:  
Pred. No.: 8,63e-168 Length: 3974  
Score: 3320.50 Matches: 645  
Percent Similarity: 79.56% Conservative: 48  
Best Local Similarity: 74.05% Mismatches: 65  
Query Match: 72.72% Indels: 113  
DB: 10 Gaps: 12

US-09-720-086-8 (1-853) x AF151976 (1-3974)

Qy 1 MetLysGlyAspThrArgHisLeuAsnGlyGluGluAspAlaGlyArgGluAspSer 20  
Db 297 ATGAAGGGAGAGCAGCAGACATCTGAATGAAGAAGAGGGTGCACGGGTATGAGAGGTGC 356  
Qy 21 IleLeuValAsnGlyAlaCysSerAspGlnSerSerAsp-----SerPro 35  
Db 357 ATTATCGTTAATGGGAACCTTCAGTGACCATCTCTCAGACACGAAAGGATGCTCCCTCACCC 416  
Qy 36 ProIleLeuGluAlaIle-----ArgThrProGluIleArgGlyArgArg 50  
Db 417 CCAGTCTTGGAGGCAATCTGCACAGAGCCAGTCTGCACACACGAGACAGAGCGCGCAGG 476  
Qy 51 SerSerArgLeuSerLysArgGluValSerSerIleLeuSerThrGlnAspLeu 70  
Db 477 TCAAGCTCCCGCTGTCTAAGAGGAGGTCTCAGCGCTTCTGAATTACACGACGACATG 536  
Qy 71 ThrGlyAspGlyAspGly-----GluAspGlyAspGlySerAspThrProVal 86  
Db 537 ACAGAGATGGAGACAGAGATGATGAATAGATGATGGGAATGCTCTGATATT--CTA 593  
Qy 87 MetProLysLeuPheArgGlu-----ThrArgThrArgSerGluSerProAlaVal 103  
Db 594 ATGCCAAAGCTCACCGGTGAGACCAAGGACACGAGGACGGCTCTGAAGCGCGCTGTC 653  
Qy 104 ArgThrArgAsnAsnSerValSerSerArgGluArgHisArgProSerProArgSer 123  
Db 654 CGAACCCGACATAGCAATGGGACCTCCAGCTTGAGAGGCAAGAGCGCTCCCGCAATC 713  
Qy 124 ThrArgGlyArgGlnGlyArgHisValAspGluSerProValGluPheProAlaThr 143  
Db 714 ACCCGAGGTGGAGGCGCCGCCACCATGTCCAGAGTACCTGTGGAGTTTCCGGCTACC 773  
Qy 144 ArgSerLeuArgArgAlaThrAlaSerAlaGlyThrProThrProSerProSer 163  
Db 774 AGGTCTCGGAGACGTCCAGCATCGTCTTCAGCAAGCACGCGCATGCTATCCCTGCCACG 833  
Qy 164 SerTyLeuThrIleAspLeuThrAspThrGluAspThrHisGlyThrProGlnSer 183  
Db 834 -----GTCGACTTCATGGAGAGAGTG-----ACACCTAAGAGC 866  
Qy 184 SerSerThrProTyAlaArgLeuAlaGlnAspSerGlnGlnGlyMetGluSerPro 203  
Db 867 GTCAGTACCCCATCAGTTGACCTTGAGCCAGGATGGAGATCAGGAGGTATGGATACCACA 926  
Qy 204 GlnValGluAlaAspSerGlyAspGlyAspSerSerGluTyGlnAspGlyGluPhe 223  
Db 927 CAGGTGGATGCAGAGACATATATGAGACAGCAGCATATCAGGATGATAAAGAGTTT 986  
Qy 224 GlyIleGlyAspLeuValTrpGlyIleLeuGlyPheSerTrpTrpProAlaMetVal 243  
Db 987 GGAATAGGTGACCTCGTGTGGGAAAGATCAAGGCTTCTCTGTGGCGCTGCCATGGTG 1046  
Qy 244 ValSerTrpLysAlaThrSerLysArgGlnAlaMetSerGlyMetArgTrpValGlnTrp 263  
Db 1047 GTGCTCTGGAAGCCACCTCCAGGCGACAGCGCATGCCGGAATGCGCTGGGTACAGTGG 1106  
Qy 264 PheGlyAspGlyLysPheSerGluValSerAlaAspLysLeuValAlaLeuGlyLeuPhe 283

Db 1107 TTGGTGATGGCAAGTTTCTGAGATCTCTGCTCAAACTGGTGGCTCTGGGGCTGTTC 1166  
Qy 284 SerGlnHisPheAsnLeuAlaThrPheAsnLysLeuValSerTyArgLysAlaMetTy 303  
Db 1167 AGCCAGCAGCTTTAATCTGGCTACCTTCAATAAGCTGGTTTCTTATAGGAAGCCATGTAC 1226  
Qy 304 HisAlaLeuGluLysAlaArgValArgAlaGlyLysThrPheProSerSerProGlyAsp 323  
Db 1227 CACACTCTGAGAAAGCCAGGGTTCGAGCTGGCAAGACCTTCTCCAGCAGTCTCTGGAGAG 1286  
Qy 324 SerLeuGluAspGlnLeuLysProMetLeuGluTrpAlaHisGlyGlyPheLysProThr 343  
Db 1287 TCACTGGAGGACAGCTGAAGCCCATCTGGAGTGGGCCACGGTGGCTTCAAGCCTACT 1346  
Qy 344 GlyIleGluGlyLeuLysProAsnAsnThrGlnProValValAsnLysSerLysValArg 363  
Db 1347 GGATCGAGGGCTCAAAACCCCAACAGACGACCC----- 1382  
Qy 364 ArgAlaGlySerArgLysLeuGluSerArgLysTyArgLysThrGluAsnLysThrArgArgThr 383  
Db 1383 -----GAGAAACAAAAGTCGAAGACGCACA 1406  
Qy 384 AlaAspAspSerAlaThrSerAspTyCysProAlaProLysArgLeuLysThrAsnCys 403  
Db 1407 ACCAATGACTCTGCTGCTCTGAG---TCCCCCCCCACCAAGCGCTCAAGACAAATAGC 1463  
Qy 404 TyrAsnAsnGlyLysAspArgGlyAspGluAspGlnSerArgGluGlnMetAlaSerAsp 423  
Db 1464 TAT---GGCGGGAAGGACCCGAGGAGGATGAGGAGGACCGAGACCGATGGCTTCTGAA 1520  
Qy 424 ValAlaAsnAsnLysSerSerLeuGluAspGlyCysLeuSerCysGlyArgLysAsnPro 443  
Db 1521 GTCACCAACAAAGGCAATCTGGAAGACCGCTGTTTGTCTCTGTGGAAGAAGAACCCCT 1580  
Qy 444 ValSerPheHisProLeuPheGlyGlyLeuGlyGlnThrCysArgAspArgPheLeu 463  
Db 1581 GTGCTCTTCACCCCTCTTTGAGGGTGGGCTCTGTACAGATTGCGGGATCGCTTCTCTA 1640  
Qy 464 GluLeuPheTyMetTyArgAspAspGlyTyArgGlnSerTyCysThrValCysCysGlu 483  
Db 1641 GAGCTCTTCTACATGATGATGAGGACGGCTATCAGTCTACTGCACCGTGTGCTGTAG 1700  
Qy 484 GlyArgGluLeuLeuLeuCysSerAsnThrSerCysCysArgCysPheCysValGluCys 503  
Db 1701 GGCCGTGAAGTGTGCTGTGTCAGTAACACAAAGCTGCTGCAGATGCTTCTGTGTGAGTGT 1760  
Qy 504 LeuGluValLeuValGlyThrGlyThrAlaAlaGluAlaLysLeuGlnGluProThrSer 523  
Db 1761 CTGGAGGTGCTGTGGCGCAGGCACAGCTGAGGATGCCAAGCTGCAGGAACCCCTGGAGC 1820  
Qy 524 CysTyMetCysLeuProGlnArgCysHisGlyValLeuArgArgArgLysAspTrpAsn 543  
Db 1821 TGCTATATGTGCTCTCCCTCAGCGCTGCATGGGCTCTCCGACGCGAAGAGATTGGAAC 1880  
Qy 544 ValArgLeuGlnAlaPhePheThrSerAspThrGlyLeu---GluTyCLeuAlaProLys 562  
Db 1881 ATGCGCTGCGAGACTTCTTCACTACTGATCTGACCTGGAAGAAATTTGAGCCACCCCAAG 1940  
Qy 563 LeuTyProAlaIleProAlaAlaArgArgProIleArgValLeuSerLeuPheAsp 582  
Db 1941 TTGTACCCAGCAATCTCTGCGAGCAAAAGGAGGCCATTAGAGTCTCTGTCTGTGTGAT 2000  
Qy 583 GlyIleAlaThrGlyTyLeuValLeuLysGluLeuGlyIleLysValGlyLysTyVal 602  
Db 2001 GGAATTGCAACGGGTACTTGTGTCTCAAGGAGTGGGTATTAAGTGGAAAGATACATT 2060  
Qy 603 AlaSerGluValCysGluGluSerIleAlaValGlyThrValLysHisGluGlyAsnIle 622  
Db 2061 GCTCTCGAAGTCTGTGCAGAGTCCATCGCTGTGGAACTGTTAAGCATGAAGCCGAGATC 2120  
Qy 623 LysTyValAsnAspValArgAsnIleThrLysLysAsnIleGluGluTrpGlyProPhe 642



```

Db 2121 AATATGTCATGACGTCGCGAAATACACAGAAATATTTGAAGAGTGGGCCCGTTTC 2180
QY 643 AspleuValIleGlyGlySerProCysAsnAspleuSerAsnValAsnProAlaArgLys 662
Db 2181 GACTTGATGATGTGAGAGCCCATGATGATCTCTTAAGTCATCTGCCCGCAA 2240
QY 663 GlyLeuTyrGluGlyThrGlyArgLeuPheGluPheTyrHisLeuLeuAsnTyrSer 682
Db 2241 GGTATATATGAGGGGACAGAGAGCTCTTCTTGAAGTTTACACTTGTCTGAATATATACC 2300
QY 683 ArgProLysGluGlyAspAspArgProPheTyrMetPheGluAsnValValAlaMet 702
Db 2301 CGCCCAAGAGGAGGCGACAAACGTCATCTCTTGATGTTGAGAAATGTTGGCCATG 2360
QY 703 LysValGlyAspLysArgAspIleSerArgPheLeuGluCysAsnProValMetIleAsp 722
Db 723 361 AAAGTGAATGACAAAGAAACATCTCAAGATTCCTGGCATGTAAACCAAGTATGATCGAT 2420
QY 723 AlaIleLysValSerAlaAlaHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMet 742
Db 2421 GCCATCAAGGTGTCTGTCTGCACAGGCCGCGTACTTCTGGGTAACTTACCCGGAATG 2480
QY 743 AsnArgProValIleAlaSerLysAsnAspLysLeuGluLeuGlnAspCysLeuGluTyr 762
Db 2481 AAC----- 2483
QY 763 AsnArgIleAlaLysLeuLysValGlnThrIleThrLysSerAsnSerIleLys 782
Db 2483 ----- 2483
QY 783 GlnGlyLysAsnGlnLeuPheProValValMetAsnGlyLysGluAspValLeuTyrCys 802
Db 2483 ----- 2483
QY 803 ThrGluLeuGluLysGlyPheGlyPheProValHisTyrThrAspValSerAsnMetGly 822
Db 2484 -----AGATCTTGGCTTCCTGCTCACTACACGACGTCGTCACACATGGGC 2531
QY 823 ArgGlyAlaArgGlnLysLeuLeuGlyArgSerTrpSerValProValIleArgHisLeu 842
Db 2532 CGCGCGCGCGCGTGAAGAGCTGTGGCAGGTCTCGAGTGTACCGGTCAACAGACACCTG 2591
QY 843 PheAlaProLeuLysAspTyrPheAlaCysGlu 853
Db 2592 TTTGCCCTTGAAGAGACTTGTGCTGTGAA 2624

```

Search completed: November 22, 2002, 10:59:08  
 Job time : 4955.31 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 22, 2002, 03:58:32 ; Search time 5196.49 Seconds

(without alignments)  
5107.633 Million cell updates/sec

Title: US-09-720-086-7  
Sequence: 1 MPAMPSSSGPGRSSSAARE.....WSVPVIRHLPFLKEYFACV 912

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool/US09720086/runat.18112002.092832.22195/app.query.fasta.1.4252  
-DB=genEmbl -QPM=fastap -SUPFIX=rge -MINMATCH=0.1 -LOOFC=0 -LOOPEXT=0  
-UNITS=bits -STAR=1 -END=1 -MATRIX=biosum62 -TRANS=human4.0.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=prc -NORMext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09720086.cgn.1.1.1308@runat.18112002.092832.22195 -NCPU=6 -ICPU=3  
-NO\_XUPXY -NO\_MAP -LARGEQUERY -NEG SCORES=0 -WAIT -LONGLOG -DEV\_TIMEOUT=120  
-WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELext=7

Database :

GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_hlg:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_scs:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_scs:\*  
28: em\_un:\*

29: em\_vl:\*  
30: em\_hlg\_hum:\*  
31: em\_hlg\_inv:\*  
32: em\_hlg\_other:\*  
33: em\_hlg\_mus:\*  
34: em\_hlg\_pin:\*  
35: em\_hlg\_rtd:\*  
36: em\_hlg\_mam:\*  
37: em\_hlg\_vrt:\*  
38: em\_sy:\*  
39: em\_hngo\_hum:\*  
40: em\_hngo\_mus:\*  
41: em\_hngo\_other:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4935	100.0	3005	9 AF067972	AF067972 Homo sapi
2	4935	100.0	4258	9 AF331856	AF331856 Homo sapi
3	4740.5	96.1	4192	10 AF068625	AF068625 Mus muscu
4	4737.5	96.0	4094	10 BC007466	BC007466 Mus muscu
5	2541.5	51.5	123936	2 AC120824	AC120824 Rattus no
6	2413	48.9	2191	9 AK025230	AK025230 Homo sapi
7	2228.5	45.2	4163	10 AF151974	AF151974 Mus muscu
8	2227.5	45.1	166651	9 AC009474	AC009474 Homo sapi
9	2226	45.1	4135	10 AF068627	AF068627 Mus muscu
10	2225	45.1	4278	10 AF151970	AF151970 Mus muscu
11	2223.5	45.1	4223	10 AF151973	AF151973 Mus muscu
12	2221	45.0	4195	10 AF068626	AF068626 Mus muscu
13	2220	45.0	4338	10 AF151969	AF151969 Mus muscu
14	2211	44.8	4335	9 AF331857	AF331857 Homo sapi
15	2203	44.6	4267	9 AF176228	AF176228 Homo sapi
16	2199.5	44.5	176697	2 AC112586	AC112586 Rattus no
17	2196	44.5	4145	9 AF156488	AF156488 Homo sapi
18	1956	39.6	3974	10 AF151976	AF151976 Mus muscu
19	1953.5	39.6	3946	10 AF068628	AF068628 Mus muscu
20	1952.5	39.6	4089	10 AF151971	AF151971 Mus muscu
21	1951	39.5	4034	10 AF151975	AF151975 Mus muscu
22	1948.5	39.5	4006	10 AF078427	AF078427 Mus muscu
23	1947.5	39.5	4149	10 AF151972	AF151972 Mus muscu
24	1917.5	38.9	3897	9 AF156487	AF156487 Homo sapi
25	1806.5	36.6	197652	9 AC012074	AC012074 Homo sapi
26	1784.5	36.2	204534	2 AC118195	AC118195 Mus muscu
27	1784.5	36.2	208874	2 AC111092	AC111092 Mus muscu
28	1769	35.8	2848	5 AF135438	AF135438 Dario rex
29	1624.5	32.9	3017	9 AK001191	AK001191 Homo sapi
30	1504	30.5	1758	9 BC018214	BC018214 Homo sapi
31	1409.5	28.6	210269	2 AC116459	AC116459 Mus muscu
32	1142.5	23.2	185413	2 AC119009	AC119009 Rattus no
33	1006.5	20.4	119630	2 AC112040	AC112040 Rattus no
34	996.5	20.2	167568	2 AC111734	AC111734 Rattus no
35	994	20.1	2127	9 AF129267	AF129267 Homo sapi
36	872.5	17.7	2057	9 AF129268	AF129268 Homo sapi
37	785	15.9	2008	9 AF129269	AF129269 Homo sapi
38	748	15.2	2077	6 AR129189	AR129189 Sequence
39	728.5	14.8	119630	2 AC112040	AC112040 Rattus no
40	693.5	14.1	1397	9 BC002560	BC002560 Homo sapi
41	689	14.0	1705	9 AF194032	AF194032 Homo sapi
42	643.5	13.0	1589	10 AF220524	AF220524 Mus muscu
43	643.5	13.0	1645	2 AF220524	AF220524 Mus muscu
44	562.5	11.4	244329	2 AC107644	AC107644 Mus muscu
45	465	9.4	117264	2 AC103454	AC103454 Rattus no

RESULT 1

#### ALIGNMENTS

AF067972 3005 bp mRNA linear PRI 12-FEB-2001  
 LOCUS Homo sapiens DNA cytosine methyltransferase 3 alpha (DNMT3A) mRNA,  
 complete cds.  
 AF067972  
 VERSION AF067972.2 GI:12746531  
 KEYWORDS  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 3005)  
 AUTHORS Xie, S., Wang, Z., Okano, M., Nogami, M., Li, Y., He, W.W., Okumura, K.  
 and Li, E.  
 TITLE Cloning, expression and chromosome locations of the human DNMT3  
 gene family  
 JOURNAL Gene 236 (1), 87-95 (1999)  
 MEDLINE 99365304  
 PUBMED 10433969  
 REFERENCE 2 (bases 1 to 3005)  
 AUTHORS Xie, S., Okano, M. and Li, E.  
 TITLE Direct Submission  
 JOURNAL Submitted (25-MAY-1998) CVRC, Mass. General Hospital, 149 13th  
 Street, Charlestown, MA 02129, USA  
 REFERENCE 3 (bases 1 to 3005)  
 AUTHORS Hata, K., Shirohzu, H., Sasaki, H. and En, L.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-FEB-2001) CVRC, Mass. General Hospital, 149 13th  
 Street, Charlestown, MA 02129, USA  
 REMARK Sequence update by submitter  
 COMMENT On Feb 12, 2001 this sequence version replaced gi:4927369.  
 FEATURES  
 Location/Qualifiers  
 1..3005  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 /map="2p23; between D2S171 and D2S174; 48-50 cM"  
 1..3005  
 /gene="DNMT3A"  
 230..2968  
 /gene="DNMT3A"  
 /function="de novo DNA cytosine methyltransferase"  
 /codon\_start=1  
 /product="DNA cytosine methyltransferase 3 alpha"  
 /protein\_id="A4D3084.2"  
 /db\_xref="GI:12746532"  
 /translation="MPAMPSSGFGDTSSAAAREEDRKGEEPRKBERQBPSTT  
 ARKVRGRKRKHPPVSGDTPKDPVAVISKPSMAQDSASELLPNGLDKRSEPPQE  
 EGSPPAGQGGAPAGEGAETLPEASRAVENGCCTPKEGRGAPAEAGKEKQETNIES  
 MKMGSRGLRGLGWESLRLORPMPRLTFOAGDPYVYSKRKDEWLARKREAEKKA  
 KVIACMNAVEENQGESOKVEASPPAVQPTDPASTVATTPEPVSGDAGDKNATK  
 AGDDPEYEDGEGFGLGELVNGKLRGFSWPGRIKSVMTGRSRAEGSTFVWVFGDG  
 KPSVVCVKLMLPSSFCSAFHQNTYKQPMYKAIIEVLOVASSRAGKLPVCHDSDE  
 SDTAKAVEONKPMIEWALGFGPSPGKGLPEPPEEKNPYKEVITDMWVPEAAAYAP  
 PPPAKPRKSTAEKPKVKEIIDEETRELVYVRQCRNIEDICISGSLNVTLEHPL  
 FVGMCKQCKNCFLECAQYDDGYQSYCTICCGREVLMCGNNCCRCFCVECDLL  
 VGPQAQAAIKEDPWNQCMGHKGYLLLRREDWPSRLQWFFANNHDBQFDPKVPY  
 PVPABKRPVRLSLFDGIATGLLVKDLGIQVDRYIASVECDSTVGMVVRHCKIM  
 YVGDVRSVTOKHIOEWPFDLVIGGSPCNDLSIVNPAKGLYEGTGLRFFEFYRLHLD  
 APRKGGDRPPFWLENVVMGVSQDKDJSRFLSNPMDIAKEVSAHARAPWGNL  
 PGMRLPLASTVNDKLEQLCEHGRIAKSKVPTITRSNSIKQKQDHPFVFNVEKE  
 DMLWCTEMERVFVPHYITDVNSMLAKRLQRLGRSWSVPIRHLFAPLKEYFACY"

gene

CDS

BASE COUNT 674 a 846 c 978 g 507 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.67e-188 Length: 3005  
 Score: 4935.00 Matches: 912  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0

US-09-720-086-7 (1-912) x AF067972 (1-3005)  
 QY 1 MetProAlaMetProSerSerGlyProGlyAspThrSerSerSerAlaAlaGluArgGlu 20  
 DB 230 ATGCCCGCCATGCTCCAGCGGCCCGCGGGACACACAGCAGCTCTGCTCGGAGCGGGAG 289  
 QY 21 GluAspArgLysAspGlyGluGluGluGluGluGluGluGluGluGluGluGlu 40  
 DB 290 GAGGACCGAAGAGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 349  
 QY 41 ProSerThrThrAlaArgLysValGlyArgProGlyArgLysArgLysArgLysProVal 60  
 DB 350 CCCAGCACACGCGACGGAAGGTGGGCGGCTGGAGGAAGCGCAAGCACCCCGGTG 409  
 QY 61 GluSerClyAspThrProLysAspProAlaValIleSerLysSerProSerMetAlaGln 80  
 DB 410 GAAAGCGGTGACACGCCCAAGGACCTTGGGTGATCTCCAAGTCCCCATCCATGCCCGAG 469  
 QY 81 AspSerGlyAlaSerGluLeuLeuProAsnGlyAspLeuGluLysArgSerGluProGln 100  
 DB 470 GACTCAGGCGCCTCAGAGCTATTACCCCAATGGGACTTGGAGAGCGGAGTGAGCCCCAG 529  
 QY 101 ProGluGluGlySerProAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 120  
 DB 530 CAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 589  
 QY 121 AlaAlaGluThrLeuProGluAlaSerArgAlaValGluAsnGlyCysCysThrProLys 140  
 DB 590 GCAGCTGAGACCTGCTGAAAGCTTCAAGAGCAGTGGAAAAATGGCTGCTGCAACCCCAAG 649  
 QY 141 GluGlyArgGlyAlaProAlaGluAlaGlyLysGluGlnLysGluThrAsnIleGluSer 160  
 DB 650 GAGGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 709  
 QY 161 MetLysMetGluGlySerArgGlyArgLeuArgGlyGlyGlyGlyGlyGlyGlyGly 180  
 DB 710 ATGAAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 769  
 QY 181 ArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTyrTyrIleSerLys 200  
 DB 770 CGTCAGCGGCCATGCGGAGGCTCACCTTCCAGCGGGGACCCCTACTACATCAGCAAG 829  
 QY 201 ArgLysArgAspGluThrLeuAlaArgTriLysArgGluAlaGluLysLysAlaLysVal 220  
 DB 830 CGCAAGCGGAGCAGGTGGCTGGCAGCTGGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 889  
 QY 221 IleAlaGlyMetAsnAlaValGluGluGluGluGluGluGluGluGluGluGluGlu 240  
 DB 890 ATTGCAGGAATGAATGCTGTGGAAGAAAAACAGGGGCGCGGGGAGTCTCAGAAAGGTGGAG 949  
 QY 241 GluAlaSerProProAlaValGluGlnProThrAspProAlaSerProThrValAlaThr 260  
 DB 950 GAGGCGAGCCCTCTGCTGTGACAGACCCCTACACCCCGCATCCCCCTGTGCTACCTACC 1009  
 QY 261 ThrProGluProValGlySerAspAlaGlyAspLysAsnAlaThrLysAlaGlyAspAsp 280  
 DB 1010 ACSCCTGAGCCCCGTGGGTCCTGATGCTGGGACCAAGAAATGCCCAAGAGCGGCGATGAC 1069  
 QY 281 GluProGluThrGluAspGlyArgGlyPheGlyLysGlyGluLeuValTropGlyLysLeu 300  
 DB 1070 GAGCCAGAGTACAGAGACGCGCGGGGCTTGGCATTTGGGAGCTGGTGTGGGGGAAACTG 1129  
 QY 301 ArgGlyPheSerThrProGlyArgIleValSerThrProMetThrGlyArgSerArg 320  
 DB 1130 CGGGGCTTCTCTGTTGGCGAGCGCGCATTTGTCTTGTGGATGACGGCGCGGAGCCGA 1189  
 QY 321 AlaAlaGluGlyThrArgTriValMetTriPheGlyAspGlyLysPheSerValValCys 340  
 DB 1190 GCAGCTGAAGGACCCCGCTGGTGTGTTGGTTCGAGACGCGCAAAATCTCAGTGGTGTGT 1249  
 QY 341 ValGluLysLeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsn 360  
 DB 1250 GTTGAGAGCTGATCGCGCTGAGCTCGTTTTTGCAGTGGTTCACCCAGGCCACGTACAAC 1309

```

Oy 361 LysGlnProMetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAla 380
Db 1310 AAGGAGCCCATGTACCGCAAGCCATCTACGAGGTCTTGCGAGGTGGCCGACGCCGCGCG
Oy 381 GlyLysLeuPheProValCysHisAspSerAspGluSerAspThrAlaLysAlaValGlu 400
Db 1370 GGGAGAGCTTCCGGGTGTGCGACAGCGAGTGAAGTGCACCTGCCAAGCGCTGGAG
Oy 401 ValGlnAsnLysProMetIleGluTyrAlaLeuGlyGlyPheGlnProSerGlyProLys 420
Db 1430 GTGAGAAACAAACCCCATGATTGAATGGGCCCTGGGGGCTTCCAGCTTCTGGCCCTAAG
Oy 421 GlyLeuGluProProGluGluGluLysAsnProTyrTyrLysGluValTyrThrAspMetTyr 440
Db 1490 GGCCTGGAGCCACCAAGAGAGAGAGAAATCCCTACAAAGAGTGTACCGGACATGTGG
Oy 41 ValGluProGluAlaAlaAlaTyrAlaProProProAlaLysLysProArgLysSer 460
Db 1550 GTGGAACCTGAGGACCTGCTTACGACCACTCCACACGCAAAAGCCCGGAAAGAGC
Oy 461 ThrAlaGluLysProLysValLysGluIleIleAspGluArgThrArgGluArgLysVal 480
Db 1610 ACACCGGAGAAAGCCCAAGGTCAAGAGATTTGATGAGCGCACAGAGAGCGGCTGGTG
Oy 481 TyrGluValArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeu 500
Db 1670 TACAGAGGTGGCGAAGAGTGGCGGAACATTGAGAGCATCTGCATCTCTGTGGAGCTTC
Oy 501 AsnValThrLeuGluHisProLeuPheValGlyGlyMetCysGlnAsnCysLysAspCys 520
Db 1730 AATGTTACCTGAGAACACCCCTCTTGTGGAGAAATGTGCCAAACTGGCAAGAACTGC
Oy 521 PheLeuGluCysAlaIleTyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrIleCys 540
Db 1790 TTTTGGAGTGTGCGTACCACTACACAGCGAGCTACAGCTTCTACTGCACCATCTGC
Oy 541 CysGlyGlyArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysVal 560
Db 1850 TGTGGGGCGCGAGAGGTCTCATGTGCGGAAACAACTCTGCAAGGTGCTTTTGGCTG
Oy 561 GluCysValAspLeuLeuValGlyProGlyAlaAlaGlnAlaAlaIleLysGluAspPro 580
Db 1910 GAGGTGTGAGCTCTTGTGTGGGCGCGGGGCTGCCAGGAGCACTTAAAGAAAGACCCC
Oy 581 TrpAsnCysTyrMetCysGlyHisLysGlyTyrTyrGlyLeuLeuArgArgGluAsp 600
Db 1970 TGGAACTGCTACTGTGCGGACCAAGGTACTTACGGGCTGTGCGGCGCGAGAGGAC
Oy 601 TrpProSerArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProPro 620
Db 2030 TGGCCCTCCCGGCTCCAGATGTTCTTCGCTAATTAACCAAGACAGAAATTTGACCTTCCA
Oy 621 LysValTyrProProValProAlaGluLysArgLysProIleArgValLeuSerLeuPhe 640
Db 2090 AAGGTTTACCACTGCTGCCAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
Oy 641 AspGlyIleAlaIleThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyr 660
Db 2150 GATGGAATGCTACAGGAGCTCTGCTGCTGAAGAGCTTGGGCACTTCAAGTGGAGCGCTAC
Oy 661 IleAlaSerGluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLys 680
Db 2210 ATTGCTCGGAGGTGTGTGAAGACTTCATCAGGTGGGCACTGTGGGCAAGGGGAGAG
Oy 681 IleMetTyrValGlyAspValArgSerValThrGlnLysHisIleGlnGluTyrGlyPro 700
Db 2270 ATCATGTAGCTGGGGAGCGTCCGAGCGTCAACAGAGCATATCAGAGATGGGGCCCA
Oy 701 PheAspLeuValIleGlyGlySerProCysAsnAspLeuSerIleValAsnProAlaArg 720
Db 2330 TTCGATCTGTGATTTGGGGGAGAGTCCCTGCAATGATGACTCTCATGTCACCCCTGCTGC

```

---

```

Oy 721 LysGlyLeuTyrGluGlyTyrThrGlyArgLeuPheGluPheTyrArgLeuLeuHisAsp 740
Db 2390 AAGGGCTCTACAGAGGGCACTGGCGGCTCTTCTTGTAGTTCTACCGCTCTCATGAT
Oy 741 AlaArgProLysGluGlyAspAspArgProPhePheTyrLeuPheGluAsnValAla 760
Db 2450 GCCGGCCCAAGAGAGGAGATGATGCCCTTCTTCTTGTGCTCTTTGAGATGTGTGGCC
Oy 761 MetGlyValSerAspLysArgAspIleSerArgPheLeuGluSerAspProValMetIle 780
Db 2510 ATGGCGCTTAGTACAGAGAGGACATCTCGCATTTCTCAGTCCAACTGTGTGATGATT
Oy 781 AspAlaLysGluValSerAlaAlaHisArgAlaArgTyrPheTyrPheTyrGlyAsnLeuProGly 800
Db 2570 GATCCCAAGAAAGTGTCAAGTGTGCACAGGCGCCGCTACTTCTGGGGTAACTTCCCGGT
Oy 801 MetAsnArgProLeuAlaSerThrValAsnAspLysLeuGlnGluCysLeuGlu 820
Db 2630 ATGAAACAGGCCCTTGGCATCTCAGCTGTGATGATTAAGCTGAGCTGCAGAGAGTGTGGAG
Oy 821 HisGlyArgIleAlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIle 840
Db 2690 CATGGCAGATAGCCAAAGTTCAAGCAAGTGAAGACCATTTCTACGAGTCAAACTCCATA
Oy 841 LysGlnGlyLysAspGlnHisPheProValPheMetAsnGluLysGluAspIleLeuTyr 860
Db 2750 AAGCAGGGCAAAAGACCAAGCATTTCTGTCTTCATGATGAGAAAGAGACATCTTATGG
Oy 861 CysThrGluMetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMet 880
Db 2810 TGCACTGAATGGAAGAGGATTTGGTTTCCAGTCCACTATGACTGACTCTCCACATG
Oy 881 SerArgLeuAlaArgGlnArgLeuLeuGlyArgSerTyrPheSerValProValIleArgHis 900
Db 2870 AGCGGCTTGGCGAGAGAGACTCTGCGCGGTCAATGAGCGGTGCCAGTCACTCCGCGCAC
Oy 901 LeuPheAlaProLeuLysGluTyrPheAlaCysVal 912
Db 2930 CTCTTCGCTCCGCTGAAGAGATATTTCGTGTGTG 2965

```

**RESULT 2**  
**LOCUS** AF331856  
**DEFINITION** Homo sapiens DNA cytosine methyltransferase 3 alpha (DNMT3A) mRNA,  
 complete cds.  
**ACCESSION** AF331856  
**VERSION** AF331856.1 GI:18033252  
**KEYWORDS**  
**SOURCE** Homo sapiens.  
**ORGANISM** Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
**REFERENCE**  
**AUTHORS** Ni,J., Pradhan,S. and Roberts,R.J.  
**TITLE** Cloning, expression and characterization of human DNMT3 genes  
**JOURNAL** Unpublished  
**REFERENCE** 2 (bases 1 to 4258)  
**AUTHORS** Ni,J., Pradhan,S. and Roberts,R.J.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (22-DEC-2000) New England Biolabs, 32 Tozer Road,  
 Beverly, MA 01915, USA  
**FEATURES**  
**source**  
 1..4258  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="2"  
 1..4258  
 /gene="DNMT3A"  
 /gene="DNMT3A"  
 227..2956  
 /gene="DNMT3A"  
 /codon\_start=1  
 /product="DNA cytosine methyltransferase 3 alpha"  
 /protein\_id="AA157039.1"

```

/db_xref="GI:18033253"
/translation="MPSSPGDGTSSAAAREEDRKDEGEQEEPRGKEERPEPSTARK
VGRPKRKHPPVEGDTTPKPAVIAISKSPMAQDSGASSELLPNGDLKESRPEQEEGS
PAGGOKGAPAGEGAETLPASRAVENGCCTPKRGRGAPAEAGKQETNIEISMVM
EGSRGLRGGLGWESSLRORPRLTFOAGDPVYISKRRDEMLARKREAEKKAVI
AGMVAENOGGEGESOKVEASPPAVQPTDPAFTVATTPPEVGSAGDKNATKAGD
DEVEDGPGFGIGELVWGLKRGFSWHPGRIVSNWMTGRSRAAEGTRVWVFGDKES
VVCSEKMLPSLSCFAFHQATYINKMYRKAIYEVLOVASRAGLTFVCHDSDESDT
AKAVEQNKPMLEWALGGFOPSGKLEPPEBEKNPYEYITDMWVPEAAAYAPPPP
AKKPRKSTAEKPKVKEIIDERTRELYEVROKRNIEDICI SCGSINVTLEHPLFVG
GMQCNKCFLECAVQDDGYOSYCTICCGREVLMCGNNCCRCFCVCEVDLLVGP
GAAQAAIKEDPNWCMCHKGTGLRRREDWPSRLQMFANNDHDFDPKVPVPPV
AEKRPVRLSLFDGATGLLVKLDLGIQVDRYIASEVCEDSITVMVRHOGKIMYVG
DVRSTOKHIQEMGPDFLVIGGSPCNDLSIVNPARGLYEGTGLRLEFEYRLLDARP
KEGDDRPBPMLEFNUVAMGVSDKRDLSRFLESNPVMDAKEVSAARHARVEMGNLPGM
WRPLASTVNDKLEQECLEHGRITAKSKVRTITRSNLIKQKDOHPFVPMNEKDL
NCTERWVPGFVHYDVSNMRLAKRQLLRGSVSPVIRHLFAPLKEFYACV"
BASE COUNT 1062 a 1089 c 1257 g 850 t
ORIGIN

```

Alignment Scores:	2,31e-188	Length:	4258
Pred. No.:	4935.00	Matches:	912
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	9	Gaps:	0
US-09-720-086-7 (1-912) x AF331856 (1-4258)			
Qy	1	MetProAlaMetProSerSerGlyProGlyAspThrSerSerSerAlaAlaGluArgGlu	20
Db	218	ATGCCCGCCATGCCCTCCAGCGGCCCGGGACACACAGCAGCTCTGCTGCGGAGCGGAG	277
Qy	21	GluAspArgLysAspGlyGluGluGluGluGluProArgGlyLysGluGluArgGluGlu	40
Db	278	GAGGACCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	337
Qy	41	ProSerThrThAlaAlaArgLysValGlyArgProGlyArgLysArgLysHisProProVal	60
Db	338	CCACGACCCAGCGACGAGAGGTGGGCGGCGCTGGGAGGAGGAGGAGGAGGAGGAGGAG	397
Qy	61	GluSerGlyAspThrProLysAspProAlaValIleSerLysSerProSerMetAlaGln	80
Db	398	GAAAGCGGTGACACGCGCAAGGACCTCGGTGATCTCCAAAGTCCCATCCAGCCAG	457
Qy	81	AspSerGlyAlaSerGluLeuLeuProAsnGlyAspLeuGluLysArgSerGluProGln	100
Db	458	GACTCAGGCGCCTCAGAGCTATTACCCAAATGGGAGCTTGGAGAGCGGAGTGAGCCCCAG	517
Qy	101	ProGluGluGlySerProAlaGlyGlyGlnLysGlyAlaProAlaGluGlyGlyGly	120
Db	518	CCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	577
Qy	121	AlaAlaGluThrLeuProGluAlaSerArgAlaValGluAsnGlyCysCysThrProLys	140
Db	578	GCAGCTTGACACCTCTGAGCTCAGAGCTCAGAGCTCAGAGCTCAGAGCTCAGAGCTC	637
Qy	141	GluGlyArgGlyAlaProAlaGluAlaGlyLysGluGlnLysGluThrAsnIleGluSer	160
Db	638	GAGGCGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	697
Qy	161	MetLysMetGluGlySerArgGlyArgLeuArgGlyGlyLeuGlyTyrGluSerSerLeu	180
Db	698	ATGAAATGAGGGGCTCCCGGGCGGCTCGGGGTGCTTGGGCTGGGAGTCCAGGCTC	757
Qy	181	ArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTyrTyrIleSerLys	200
Db	758	CCTCAGCGGCGGCTCAGGAGCTCAGCTTCCAGCGGGGAGCCCTCTACTCATCAGCAAG	817
Qy	201	ArgLysArgAspGluTrpLeuAlaArgTrpLysArgGluAlaGluLysLysAlaLysVal	220
Db	818	CGCAAGCGGAGCGAGTGGCTGGCAGCGCTGGAAAGGAGGAGGAGGAGGAGGAGGAGGAG	877

Qy	221	IleAlaGlyMetAsnAlaValGluGluAsnGlnGlyProGlyGluSerGlnLysValGlu	240
Db	878	ATTGACGGAATGAATGCTGTGGAAGAAAACACAGGGGCGCGGGAGTCTCAGAAAGTGGAG	937
Qy	241	GluAlaSerProProAlaValGluGlnGlnProThrAspProAlaSerProThrValAlaThr	260
Db	938	GAGGCCAGCCCTCTCTGCTGTCAGACGCCCATGACCCCGCATCCCTCTGTCGCTACCC	997
Qy	261	ThrProGluProValGlySerAspAlaGlyAspLysAsnAlaThrLysAlaGlyAspAsp	280
Db	998	ACGCTTGAGCCCGTGGGTCCTGCTGTCGAGACGCCCATGACCCCGCATCCCTCTGTCGCTACCC	1057
Qy	281	GluProGluTyrGluAspGlyArgGlyPheGlyIleGlyGluLeuValTrpGlyLysLeu	300
Db	1058	GAGCCAGAGTACAGAGACGCGCGGGGCTTGGCATTTGGGAGCTGGTGTGGGGGAAACTG	1117
Qy	301	ArgGlyPheSerTrpTrpProGlyArgIleValSerTrpTrpMetThrGlyArgSerArg	320
Db	1118	CGGGGCTTCTCTGCTGGTGGCCAGCGCCCATTTGTGCTTGTGGTGGATGACCGGCGGAGCCGA	1177
Qy	321	AlaAlaGluGlyThrArgTrpValMetTrpPheGlyAspGlyLysPheSerValValCys	340
Db	1178	GCAGCTGAAAGGACCCCGCTGGTGCATGCTGTGTCGAGACGCGCAAAATTCAGTGGTGT	1237
Qy	341	ValGluLysLeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsn	360
Db	1238	GTTGAGAGCTGATCGCGCTGAGCTGCTTTTTCAGTGGCTTCCACGAGCCACGACACAC	1297
Qy	361	LysGlnProMetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAla	380
Db	1298	AAGCAGCCCATGTACCGCAAGGCCATCTACGAGGTCTTCAGAGTGGCCAGCAGCGCGCG	1357
Qy	381	GlyLysLeuPheProValCysHisAspSerAspGluSerAspThrAlaLysAlaValGlu	400
Db	1358	GGGAAGCTGTTCCGCTGTGTCACGACGACGAGTGAAGTGAACATGCTGCCAGGCCCTGGAG	1417
Qy	401	ValGlnAsnLysProMetIleGluTrpAlaLeuGlyGlyPheGlnProSerGlyProLys	420
Db	1418	GTGCAGAAACAGCCCATGATTGAATGGGCGCTTGGGGGCTTCCAGCCTTCTGGCCCTAAG	1477
Qy	421	GlyLeuGluProProGluGluLysAsnProTyrLysGluValTyrThrAspMetTrp	440
Db	1478	GGCCTGGAGCCACAGAGAGAGAGAGAAATCCCTACAAAGAGTGTACACGACCATGTGG	1537
Qy	441	ValGluProGluAlaAlaAlaTyrAlaProProProProAlaLysLysProArgLysSer	460
Db	1538	GTGGAACCTGAGGACGCTGCTACGACACCTCCACAGCCAAAGAGCCCGGAGAGAGC	1597
Qy	461	ThrAlaGluLysProLysValLysGluIleAspGluArgThrArgGluLeuVal	480
Db	1598	ACAGCGAGAGAGCCCAAGGTCAAGGAGATATTGATGAGCGCACAAAGAGAGCGGCTGGTG	1657
Qy	481	TyrGluValArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeu	500
Db	1658	TACGAGTGGGAGAGTGGCGGACATTTGAGGACATCTGCATCTCTCTGTGGGAGCCTC	1717
Qy	501	AsnValThrLeuGluHisProLeuPheValGlyGlyMetCysGlnAsnCysLysAsnCys	520
Db	1718	AATGTTACCTTGGAACACCCCTCTTCTGTTGGAGAAATGTGCCAAACTGCAAGAACTGC	1777
Qy	521	PheLeuGluCysAlaTyrGlnTyrAspAspGlyTyrGlnSerTyrCysThrIleCys	540
Db	1778	TTTCTGGAGTGTGCTACAGTACGACGAGCGGCTACCATCTCTACTGCACCATCTGC	1837
Qy	541	CysGlyGlyArgGluValLeuMetCysGlyAsnAsnAsnCysCysAspCysPheCysVal	560
Db	1838	TGTGGGGCGCTGAGGTGCTCATGTGGGAAACAAACTGCTGCAGGTGCTTTTTCGGTG	1897
Qy	561	GluCysValAspLeuLeuValGlyProGlyAlaAlaGlnAlaIleLysGluAspPro	580
Db	1898	GAGTGTGTGACCTCTTGTGGGGCGCGGGGCTGCCAGGCGAGCATTTAAGAGAGACCCC	1957

Oy 581 TTPAANCYTYMETCYSGLYHISLYSGLYTHR-TYRGLYLEULeuAARGARGGLUASP 600  
 Db 1958 TGGAACTGCTACATGTCGGGGCACAAGGGTACTACGGGCTGCTCGCGCGCAGAGAGAC 2017  
 Oy 601 TTPProSeArXLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProPro 620  
 Db 2018 TGGCCCTCCCGGCTCCAGATGTTCTTCGCTAATTAACAAGACAGAAATTTGACCTTCCA 2077  
 Oy 621 LysValTYRProProValProAlaGluYsArgLysProIleArgValLeuSerLeuPhe 640  
 Db 2078 AAGGTTTACCACCTGTCCTCCAGCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2137  
 Oy 641 AspGlyIleAlaThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTYR 660  
 Db 2138 GATGAAATCGCTACAGGGCTCTGCTGTAAGAGACTTGAGGACTTGGGCTTCAAGTGGACGCTAC 2197  
 Oy 661 IleAlaSerGlnValCYSGlnAspSerIleThrValGlyMetValArgHisGlnGlyLys 680  
 Db 2198 ATTGCTCTCGAGGCTGTGTAGAGACTTCATCAGGTGGGCAATGTCGGGCAACAGGGGAGAG 2257  
 Oy 681 IleMetTYRValGlnYsAspValArgSerValThrGlnLysHisIleGlnGlnTYRPro 700  
 Db 2258 ATCATGTACCTCGGGGCGCTCCGACGCTCACACAGAGACATATCAGAGAGTGGGGCCCA 2317  
 Oy 701 PheAspLeuValIleGlyGlySerProCYsAsnAspLeuSerIleValAsnProAlaArg 720  
 Db 2318 TTCGATCTGCTGATTTGGGGGAGAGTCCCTGCAATGACTCTTCATGCTCAACCTGCTCGC 2377  
 Oy 721 LysGlyLeuTYRGLYGLYTYRGLYArgLeuPhePheGluPheTYRArgLeuHisAsp 740  
 Db 2378 AAGGCCCTTACAGAGGACATCGCGGCTCTTTTGAAGTCTACCGCTCTCGCATCAT 2437  
 Oy 741 AlaArgProLysGlnGlyAspAspArgProPhePheTYRLeuPheGlnAsnValAla 760  
 Db 2438 GCGGCGCCCAAGAGAGAGAGATGCGCCCTTCTTGGCTTTGAGAAATGTGTGGCC 2497  
 Oy 761 MetGlyValSerAspLysArgAspIleSerArgPheLeuGlnSerAsnProValMetIle 780  
 Db 2498 ATGGCCCTTAGTACACAGAGGACATCTCCGAGTTTCTCGAGTCCAACTGTGATGATT 2557  
 Oy 781 AspAlaLysGlnValSerAlaAlaHisArgAlaArgTYRPhETRPGLYAsnLeuProGly 800  
 Db 2558 GATGCCAAAGAAAGTGTACGCTGACACAGGCGCCGCTACTTCTGGGGTAACTTCCCGGT 2617  
 Oy 801 MetAsnArgProLeuAlaSerThrValAsnAspLysLeuGlnGlnGlnGlnGlnGln 820  
 Db 2618 ATGAACAGGCCCTTGGCATCCACTGTGATGATTAAGCTGAGAGTCAAGAGTGTGAGAG 2677  
 Oy 821 HisGlyArgIleAlaLysPheSerLysValArgThrIleThrArgSerAsnSerIle 840  
 Db 2678 CATGCCAGGATACCAAGTTCAGCAAGTAGAGCACTTACTACGAGTCAAACTCCATA 2737  
 Oy 841 LysGlnGlyLysAspGlnHisPheProValPheMetAsnGlnLysGlnAspIleLeuTYR 860  
 Db 2738 AAGCGAGGCAAGACGACATTTCTGCTTCAATGATGAGAAAGAGACATCTTATGG 2797  
 Oy 861 CysThrGlnMetGlnArgValPheGlyPheProValHisTYRThrAspValSerAsnMet 880  
 Db 2798 TGCATGGAATGAGAAAGGATTTGGTTTCCAGTCCACTATACGCTCAACATG 2857  
 Oy 881 SerArgLeuAlaArgGlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHis 900  
 Db 2858 AGCGCCTTGGCGAGAGAGACTGCTGGGCGGCTCATGAGAGGCTGCAAGTCAATCCGCCAC 2917  
 Oy 901 LeuPheAlaProLeuLysGlnTYRPhEAlaCYsVal 912  
 Db 2918 CTCTTCGCTCCGCTGAAGAGATTTTGGCTGTGTG 2953  
 RESULT 3 AF068625 4192 bp mRNA linear ROD 06-DEC-1999  
 LOCUS AF068625 4192 bp mRNA linear ROD 06-DEC-1999  
 DEFINITION Mus musculus DNA cytosine-5 methyltransferase 3A (Dnmt3a) mRNA,  
 complete cde.

ACCESSION AF068625  
 VERSION AF068625.2 GI:6449467  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Mus musculus.  
 Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE  
 AUTHORS  
 TITLE  
 Cloning and characterization of a family of novel mammalian DNA  
 (cytosine-5) methyltransferases  
 Nat. Genet. 19 (3), 219-220 (1998)  
 JOURNAL  
 MEDLINE  
 PUBMED  
 9662369  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,  
 Charlestown, MA 02129, USA  
 3 (bases 1 to 4192)  
 Okano, M., Chijiwa, T., Sasaki, H. and Li, E.  
 Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,  
 Charlestown, MA 02129, USA  
 Sequence update by submitter  
 On Nov 18, 1999 this sequence version replaced gi:3327977.  
 REMARK  
 COMMENT  
 FEATURES  
 source  
 1. .4192  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="12"  
 /map="4.0 CM"  
 1. .4192  
 /gene="Dnmt3a"  
 1. .4192  
 /gene="Dnmt3a"  
 /note="similar to EST sequences deposited in GenBank  
 accession numbers AA052791, AA111043, AA154890, AA240794,  
 AA756653, W58898, W59299, W91664, and W91665"  
 217. .2943  
 /gene="Dnmt3a"  
 /function="de novo DNA methylation"  
 /note="contains Cys-rich region"  
 /codon\_start=1  
 /product="DNA cytosine-5 methyltransferase 3A"  
 /protein\_id="AAC40177.2"  
 /db\_xref="GI:6449468"  
 /translation="MPSGPGDTSSTSSSLEREDDRKEGEEBOENRKEERQEPSSATARK  
 VGRPGRRKHPVPSSTDPKDAVTTKSSQMAQDSGSDLPNGDLEKRSBPQREGS  
 PAAQGGAGAEAGCTETPPPEASRAVENGCCTKEGASAGSGKEQKOTIESMKME  
 GSRGTRGIGWESSLRORPMRPTROAGDPYTSKRKDEWTLARKREARKARVIA  
 VMNAVENQASGESQKVEASPPAVQDPTTPATPTVATTPEPVGADGDKATKAAAD  
 EPEVEDRGFGIIGLWMLKRGFSWMPGRIVSWMTGSRRAAGTRAVMMWFGDGKFSV  
 VCEVEKLPLSSFCSAFHQATYKQPMYKAIVYLQVASSRAKGLPACHDSDESQSG  
 KAVEYONKQWLEMLGEGROPSPKGLPEPPEEKAKYKEIYCTGSGSLAVTLTLEPLFIG  
 KKPRAKSTTEPKVEIIDERTRRLYEVROKKNIEDICISGSLAVTLTLEPLFIG  
 MCQNCNCFLECAAYDDDDGYOSYCTICCGREVLQMGNNCCRCFVECVDLVGP  
 AAQAAIKEDPMNCCMGKGTGYGLRRRDEMPSLQMFPAHNDQEDDPKVPYPPVA  
 EKRPKIRVLSIPGLATGLVLDLGIQVRYLSEVEDSLTVGAWRHOKIMVGD  
 VRSVTKHIDQEWGPFDLVITGSPQNDLSLYNPAKGLVEGTGLPFEFYLHLDARPK  
 EDDDRPEFWLFENVVAMGVSDKRDLSRFLSNPVIDAKESVAHRARVFWGNLPGM  
 RLASVNDKLELOCELEHGRJAKFSKRTITTSNSIQGXQHPFVFMENKEDILW  
 CTMEBVFGRFVHTDVSNSRLARORLLRSWSVPIRHLFAPLKEYFACV"  
 BASE COUNT 1094 a 1072 c 1186 g 840 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1,266-180 Length: 4192  
 Score: 4740.50 Matches: 877  
 Percent Similarity: 97.15% Conservative: 8  
 Best Local Similarity: 96.27% Mismatches: 25  
 Query Match: 96.06% Indels: 1  
 DB: 10 Gaps: 1



US-09-720-086-7 (1-912) x AF068625 (1-4192)

QY 2 ProAlaMetProSerSerGlyProGlyAspThrSerSerSerAlaAlaGluArgGluGlu 21  
DB 211 CCAGCAATGCCCTCCAGCGGCCCGCGGGACACACAGAGCTCTCTCTGGAGCGGGAGAT 270  
QY 22 AspArgLysAspGlyGluGluGluGluGluProArgGlyLysGluGluArgGluGluPro 41  
DB 271 GATCAAAAGGAAGGAGGAACAGGAGGAGAACCGTGGCAAGGAAGAGCGCCAGAGAGCC 330  
QY 42 SerThrThrAlaArgLysValGlyArgProGlyArgLysArgLysHisProProValGlu 61  
DB 331 AGCGCCACGGCCCGGAAGGTGGGGAGCGCTGGCGGAAGCGCAAGCACCCACCGGTGAA 390  
QY 62 SerGlyAspThrProLysAspProAlaValLysSerLysSerProSerMetAlaGlnAsp 81  
DB 391 AGCACTGACACCCCAAGGACCCAGCAGTACCACCAAGTCTCAGCCCATGGCCAGGAC 450  
QY 82 SerGlyAlaSerGluLeuLeuProAsnGlyAspLeuGluLysArgSerGluProGlnPro 101  
DB 451 TCTGCCCTCAGATCTGCTACCCAATGGAGACTTGGAGAAAGCGAGTGAACCCCAACCT 510  
QY 102 GluGluGlySerProAlaGlyGlyGlnLysGlyGlyValaProAlaGluGlyGluGlyAla 121  
DB 511 GAGGAGGGGAGCCAGCTCCAGGCGAGAGGGTGGGCCCCAGCTGAAGGAGAGGA --- 567  
QY 122 AlaGluThrLeuProGluAlaSerArgAlaValGluAsnGlyCysCysThrProLysGlu 141  
DB 568 ACTGAGACCCCAAGAGCCCTCCAGAGCTGTGGAGAAATGGCTGCTGTGTGACCAAGAA 627  
QY 142 GlyArgGlyAlaProAlaGluAlaGlyLysGluGlnLysGluThrAsnLysGluSerMet 161  
DB 628 GGCGGTGGAGCTCTGAGGAGAGGGCAAGAAAGACAGAGCAGACCAACATCGAATCCATG 687  
QY 162 LysMetGluGlySerArgGlyArgLeuArgGlyGlyLeuGlyTrpGluSerSerLeuArg 181  
DB 688 AAAATGGAGGGTCTCCGGGGCCACTGCGAGGTGGCTTGGGTGGGATCCAGCTCCGT 747  
QY 182 GlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTyrTrpLysSerLysArg 201  
DB 748 CAGCGACCCATGCCAAGACTCACTTCCAGGCGAGGAGCCCTACTACATCAGCAAAACGG 807  
QY 202 LysArgAspGluTrpLeuAlaArgTrpLysArgGluAlaGluLysLysAlaLysValile 221  
DB 808 AAACGGGATGAGTGGCTGGCACCTTGGAAAGGGAGGCTGAGAAAGCAAGCCAAAGTAAAT 867  
QY 222 AlaGlyMetAsnAlaValGluAsnGlnGlyProGlyGluSerGlnLysValGluGlu 241  
DB 868 GCAGTAATGAATGCTGTGGAAGAGAACCCAGGCCCTCTGGAGAGTCTCAGAAGGTGGAG 927  
QY 242 AlaSerProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThr 261  
DB 928 GCCAGCCCTCTCTGCTGTGCAGCAGCCACCGACCCCTGCTTCTCCGACTGTGGCCACACC 987  
QY 262 ProGluProValGlySerAspAlaGlyAspLysAsnAlaThrLysAlaGlyAspAspGlu 281  
DB 988 CCTGAGCCAGTAGGAGGGATGCTGGGGACAAGAAATGCTACCAAGAGCGCCGACGATGAG 1047  
QY 282 ProGluTrpGluAspGlyArgGlyPheGlyIleGlyGluLeuValTrpGlyLysLeuArg 301  
DB 1048 CCTGAGTATGAGGATGGCGGGCTTTGGCATTTGGAGAGTGGTGTGGGGAAACTTCGG 1107  
QY 302 GlyPheSerTrpTrpProGlyArgIleValSerTrpTrpMetThrGlyArgSerArgAla 321  
DB 1108 GGCTTCTCTCTGCGCCAGGCCGAATTGTCTTGTGGATGACAGGCGCGAGCCGAGCA 1167  
QY 322 AlaGluThrThrArgTrpValMetTrpPheGlyAsnGlyLysPheSerValValCysVal 341  
DB 1168 GCTGAAGGACCTCGCTGGTCTAGTGTTCGGAGATGGCAAGTTCTCAGTGGGTGTGTG 1227  
QY 342 GluLysLeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrThrAsnLys 361

DB 1228 GAGAGGCTCATGCCGCTGAGCTCTCTCTCAGTGCATTCACACGAGCCACCTACAACAAG 1287  
QY 362 GlnProMetTrpArgLysAlaIleTrpGluValLeuGlnValAlaSerSerArgAlaGly 381  
DB 1288 CAGCCCATGTACCGCAAGCATCTACGAAGTCTCTCAGGTGGCCAGCAGCGCTGCCGGG 1347  
QY 382 LysLeuPheProValCysHisAspSerAspGluSerAspThrAlaLysAlaValGluVal 401  
DB 1348 AAGCTCTTTCCAGCTTTGCCATGACAGTGAATAAGTACAGTGGCAAGGCTGTGAAGTG 1407  
QY 402 GlnAsnLysProMetIleGluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGly 421  
DB 1408 CAGAACAAGCAGATGATTGAATGGGCCCTCGTGCCTCCAGCCCTCGGGTCTTAAGGCG 1467  
QY 422 LeuGluProProGluGluGluLysAsnProTyrLysGluValTrpThrAspMetTrpVal 441  
DB 1468 CTGGAGCCCAAGCAAGAGAGAAATCTTTACAAAGAAAGTTTACCCGACATGTGGGTG 1527  
QY 442 GluProGluAlaAlaIleAlaProProProAlaLysLysProArgLysSerThr 461  
DB 1528 GAGCCTTGAAGCAGCTGCTTACGCCCAACCCCAAGCAAGAAACCCAGAAAGAGCACA 1587  
QY 462 AlaGluLysProLysValLysGluIleAspGluArgThrArgGluArgLeuValTrp 481  
DB 1588 ACAGAGAAACCTAAGGTCAAGGAGATCATTTGATGAGGCGCAAGGGAGCGGTGTGTAT 1647  
QY 482 GluValArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsn 501  
DB 1648 GAGGTGGCCAGAAAGTGCAGAAACATCCAGGACATTTGTATCTATGTGGAGCCCTCAAT 1707  
QY 502 ValThrLeuGluHisProLeuPheValGlyGlyMetCysGlnAsnCysLysAsnCysPhe 521  
DB 1708 GTCACCCCTGGAGCACCACTCTTATTGGAGCATGTGCCAGAACTGTAAAGAACTGCTTC 1767  
QY 522 LeuGluCysAlaTrpGlnTrpAspAspAspGlyTrpGlnSerTrpCysThrIleCysCys 541  
DB 1768 TTTGGAGTGTCTTACCAAGTAAGCAGATGGGTACCACTCTATTTGCCACATCTGCTGT 1827  
QY 542 GlyArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGlu 561  
DB 1828 GGGGGCGCTGAAGTGTCTATGTGGGAAACAACAACTGCTCAGGTGCTTTTGTGTCCAG 1887  
QY 562 CysValAspLeuLeuValGlyProGlyValaAlaGlnAlaIleLysGluAspProTrp 581  
DB 1888 TGTGTGGATCTCTTGTGGGCCAGGAGCTGTCCAGGAGCCATTAAAGAAAGACCCCTGG 1947  
QY 582 AsnCysTyrMetCysGlyHisLysGlyThrTrpGlyLeuLeuArgArgGluAspTrp 601  
DB 1948 AACTGTCTACATGTCCGGGCATAAGGGCACCTATGGGTGCTGCCAAAGACGGGAAGACTGG 2007  
QY 602 ProSerArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProLys 621  
DB 2008 CCTTCTCGACTCCAGATGTTCTTTGCCAATAACCATCACCAGGAATTTGACCCCAAG 2067  
QY 622 ValTrpProProValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAsp 641  
DB 2068 GTTTACCCACCTGTGCAGCTGAGAGAGAGAGCCCATTCGCGTGTCTCTCTTTGAT 2127  
QY 642 GlyIleAlaThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTrpIle 661  
DB 2128 GGGATTTGCTACAGGGCTCTCTGTGCTGAAGACCTGGGCATCCAAAGTGGACCGCTACATT 2187  
QY 662 AlaSerGluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIle 681  
DB 2188 GCCTCCGAGGTGTGTCAGGACTCCATCACGTTGGCATGTGTCCGCGACCCAGGAAAGATC 2247  
QY 682 MetTrpValGlyAspValArgSerValThrGlnLysHisIleGlnGluTrpGlyProPhe 701  
DB 2248 ATGTAGTCCGGGACGCTCCGACGCTCACAGAAAGCATATCCAGGAGTGGGGCCATTTC 2307  
QY 702 AspLeuValIleGlySerProCysAsnAspLeuSerIleValAlaAsnProAlaArgLys 721  
DB 2308 GACCTGGTGTGGAGGAGTCCCTCCATGATGCTCTCCATTTGTCACCCCTGCCCGCAAG 2367

OY	722	GLLEuYrGLuGLiYhXrGLiYrGLeuhPePheGLiPheYrYrAGLeuuhIsaPaLa	741
Db	2368	GGACTTATAGAGGTAAGTGGCGGCTCTTCTTGAGTTCAACCGCTCTCGATGATGGC	2427
OY	742	ArgProLYsGLuGLiYAsPaSPArPProPhePheTriPLeuPheGLiAsnValAlaMet	761
Db	2428	CGGCCCCAAGAGGAGAGATGATCGCCCTTCTTGGCTCTTGAGAAATGGTGGCCATG	2487
OY	762	GLiYALSeAspLYsArGLAsPILeSeArGLPheLeuGLuSeAsnProValMetLLeAsp	781
Db	2488	GGCGTTAGTACAGAGGGAGCATCTCGGATTTCTTGAGTCAACCCCGTATGATTGAC	2547
OY	782	AlaLYsGLuValSeArLaAlaHisArGLiArGLYrPheTriPGLiYAsnLeuProGLiMet	801
Db	2548	GCCAAGAAGTGTCTGTCTGCACACAGGGCCGTTACTTCTGGGGTAACCTTCTGGCATG	2607
OY	802	AsnArGLProLeuAlaSeThrThrValAsnAspLYsLeuGLuLeuGLiYsLeuGLiHIs	821
Db	2608	AACAGCGCTTTGGCATCTCCACTGTGAATATTAAGCTGAGCTGCAAGAGTGTCTGGAGCAC	2667
OY	822	GLiYrGLiAlaLYsPheSeRLYsValArGLThrLLeThrThrArGLSeAsnSerLLeYs	841
Db	2668	GGCAGAATAGCCCAAGTTCACCAAAAGTGGAGCACCATTAACCAACAGGTCAAACTTATTAAG	2727
OY	842	GLiGLiLYsAspGLiHisPheProValPheMetAsnGLuLYsGLiAspLLeuTriPGLys	861
Db	2728	CAGCGCAAAAGACACAGCATTTCCCGCTTTCATGAACAGAAAGAGAGACATCTCGTGGTGC	2787
OY	862	ThrGLuMetGLiArGLValPheGLiYrPheProValHisYrThrAspValSeRsnMetSer	881
Db	2788	ACTGAAATGGAAAGGGGTGGTTCCTCCCGCTCCACATACAGAGCGTCTCCCAACATGAGC	2847
OY	882	ArgLeuAlaArGLiArGLiArGLeLeuGLiYrGLSerTriPSeRValProValLLeArGLiLeu	901
Db	2848	CGCTTGGCGAGGCGAGAGACTGCTGGGCCGATCGTGAGCGGTGATCCGCCACACTTC	2907
OY	902	PheAlaProLeuLYsGLiYrPheAlaCysVal	912
Db	2908	TTCCGTCGCTGAAGGAATATTTTGCCTTGATG	2940

RESULT 4				
BC007466				
LOCUS	BC007466	4094 bp	mRNA	linear
DEFINITION	Mus musculus, DNA methyltransferase 3A, clone MGC:5662			
ACCESSION	IMAGE:3492853, mRNA, complete cds.			
VERSION	BC007466			
KEYWORDS	BC007466.1	GI:13938620		
SOURCE	MGC.			
ORGANISM	house mouse.			
	Mus musculus			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1 (bases 1 to 4094)			
AUTHORS	Strausberg,R.			
TITLE	Direct Submission			
JOURNAL	Submitted (01-MAY-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
REMARK	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>			
COMMENT	Contact: MGC help desk			

Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrived by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcdepaxil@stanford.edu](mailto:mcdepaxil@stanford.edu)  
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

```

FEATURES
  source      "MAG-1000"
  location/qualifiers
    1..4094

```

CDS

```

/product="DNA methyltransferase 3A"
/protein_id="AAH07466.1"
/db_xref="GI:13938621"
/db_xref="LocustID:13435"
/translation="MPSSGGDTSSTSLEREDRKEEGBEENRKEEKEOPSATAR
VGRPERKRHHPPVESSDTPKDPAYTTKSOPMADSCSDLPNGLKESPEDEEBS
PAAGQKGAPABEGEJETPEASRAVENGCVCVKEGASAGEGKEQKPTNISMKHE
GSRGLRGGLGWESSLRORPMPLRPFQGDPLYISKKRDWLAWMKREAKAKVLVLA
VMVAVEENOAGSGEKVEEASPAVQGPDPASPTVATTPVGGDAGKNTAATDAI
EPYLEDGRGFGIGELVWGKLRGFSWMPGRIVSWMTGRSRAEBGTRVWVNFQGGK
VCEVLEPLSPFSASFQATYVNRKQPMYRAIVELVQASSPAKGLPACHDSGSDS
KAVEVONKOMIEWALGFQSGPKLEPEEENKPYKEVYTDWMEPEAAVYAPPP
KKPRKSTKEPKPKVEIIDERTEREIVYEORKNRIDICISGGLNVLTLEHPLFIG
MCQNCNKCFLBECAYQYHDDGQSYSTICCGGREGVLMGNNNCRCFCEVCVDLVGR
AAQAIXKEDPANCVMCGHKGTGLLRREDWPSRLQHPFANNHDEFDPPKYVPVP
EKRPRIYVSLFPGIATGLVLKDGIIQYDRVYASCECDSITVGMVRHQGKTVGQI
VRSVYQKHIOEMGPPDPLVITGSSPCNDLSIVNPAKGLYEGTGRLFPFYLHLDPAR
EGDDPAPFYLFEENVVAMGVSDKRDLSRPLIESNVMTAKEVSAARHARIPFNGHLPGM
RPLASTVNDKLELOECLEHGRIAKFSKVTITTRNSNICQGDQHPVPFNNKECDDILM
CTMERVAFGFPVYATDWSNMSRLRQRLILGRSWSVPVIRHIFAPLKEVACV"

```

Alignment Scores:	
Pred. No.:	1.63e-180
Score:	4737.50
Percent Similarity:	97.04%
Best Local Similarity:	96.27%
Query Match:	96.00%
DB:	10
Gaps:	1
Length:	4094
Matches:	877
Conservative:	7
Mismatches:	26
Indels:	1
.	

US-09-720-086-7 (1-912) X BC007466 (1-4094)

QY	2	ProAlaMetProSerSerGlyProGluYAspThrSerSerSerAlaAlaGluProGluGlu	21
Db	239	CCAGCAATGCTCTCCACGGGCCCCCGGGGACACACAGACGCTCTCTTGAGGGGGAGAT	298
QY	22	AspArgLysAspGlyYSgluGluGluGluGluProArgGlyYYSgluGluGluArgGlnGluPro	41
Db	299	GATGAAAGGAAGAGAGAGACGACGAGAAACCTGTGGCAAGAGAGAGCGCCACGAGACC	358
QY	42	SerThrThrAlaArgLysValGlyYAspProGlyArgLysArgLysHisProProValGlu	61
Db	359	AGCCCCACGGCCCCGGAGAGCTGGGGAGCGCTGGCCGGAGAGCGCAAGCACCCACCGGTGAA	418
QY	62	SerGlyAspThrProLysAspProAlaValIleSerLysSerProSerMetAlaGlnAsp	81
Db	419	AGCAGTGAACCCCCCAAGAGCCACGACAGTGAACCAACCAAGCTCAGCCCATGGCCACGAC	478
QY	82	SerGlyAlaSerGluLeuLeuProAsnGlyAspLeuGluLysArgSerGluProGlnPro	101
Db	479	TCTGGCCCCCTCAGATCTGCTCAACCAATGAGACCTTGAGAGAGCGGAGTGGAACCCCAACT	538
QY	102	GluGluGlySerProAlaGlyGlyGlnYSGlyValAlaProAlaGluGluGluGlyAla	121

Db 539 GAGGAGGAGCCAGCTCAGGGCAGAGGGTGGGCCCCAGCTGAGGAGAGGGA--- 595  
QY 122 AlaGluThrLeuProGluAlaSerArgAlaValGluAsnGlyCysThrProLysGlu 141  
Db 596 ACTGAGACCCACAGAGCTCCAGAGCTGTGAGAAATGGCTGCTGTGTGACCAAGGA 655  
QY 142 GlyArgGlyAlaProAlaGluAlaGlyLysGluGlnLysGluThrAsnIleGluSerMet 161  
Db 656 GGCCGTGGAGCTCTGCGAGGAGGGCAAGAAACAGAGCGGACCAACATCGAATCCATG 715  
QY 162 LysMetGluGlySerArgGlyArgLeuArgGlyGlyLeuGlyTyrGluSerSerLeuArg 181  
Db 716 AAAATGGAGGGCTCCCGGGCCGAGCTGCGAGGTGGCTTGGCTGGGAGTCCAGCTCCGT 775  
QY 182 GlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTyrTyrIleSerLysArg 201  
Db 776 CAGCAGCCATCCCAAGACTCACCTTCCAGGAGGGGACCCCTACTACATCAGCAAAACGG 835  
QY 202 LysArgAspGluTyrLeuAlaArgTyrLysArgGluAlaGluLysLysAlaLysValIle 221  
Db 836 AAACCGGATGAGTGGCTGGCACGTTGGAAAGAGGAGCTGAGAAAGAAAGCCAAAGTAAT 895  
QY 222 AlaGlyMetAsnAlaValGluGluAsnGlnGlyProGlyGluSerGlnLysValGluGlu 241  
Db 896 CGAGTAATGAATGCTGTGAAGAGAACACAGGCTCTGGAGAGTCTCAGAGGTGGAGAG 955  
QY 242 AlaSerProProAlaValGlnGlnProThrAspProAlaSerProThrValAlaThrThr 261  
Db 956 GCCAGCCCTCTCTGCTGTGCAGAGCCACAGGACCCCTCTCTCCGACTGTGGCCACCA 1015  
QY 262 ProGluProValGlySerAspAlaGlyAspLysAsnAlaThrLysAlaGlyAspAspGlu 281  
Db 1016 CCTGAGCCAGTAGGAGGGATGCTGGGGACAAAGATGCTACCAAAAGCAGCCGATGAG 1075  
QY 282 ProGluTyrGluAspGlyArgGlyPheGlyIleGlyGluLeuValTyrGlyLysLeuArg 301  
Db 1076 CCTGAGTATGAGATGGCCGGGCTTTGGCATTTGGAGAGTGGTGTGGGGGAAACTTCGG 1135  
QY 302 GlyPheSerTyrTyrProGlyArgIleValSerTyrTyrMetThrGlyArgSerArgAla 321  
Db 1136 GGCTTCTCTGGTGCCAGGCGCAATTTGTGCTTTGTGGATGACAGGCCGAGCCGAGCA 1195  
QY 322 AlaGluGlyThrArgTyrValMetTyrPheGlyAspGlyLysPheSerValValCysVal 341  
Db 1196 GCTGAAGGCACCTGGCTGGGTCAATGGTTCCGGAGATGGCAAGTTCTCAGTGGTGTGTG 1255  
QY 342 GluLysLeuMetProLeuSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLys 361  
Db 1256 GAGAAGCTCATGCCCTGAGCTCTTCTGAGTGCAATTCACAGGCCACCTCAACAAG 1315  
QY 362 GlnProMetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGly 381  
Db 1316 CAGCCCATGTACCGCAAGCCATCTACGAAGTCTCAGGTGGCCAGCAGCCGTGCCGGG 1375  
QY 382 LysLeuPheProValCysHisAspSerAspGluSerAspThrAlaLysAlaValGluVal 401  
Db 1376 AAGCTGTTTCCAGCTTGCCATGACAGTGATGAAAGTGACAGTGGCAGGCTGTGGAAAGTG 1435  
QY 402 GlnAsnLysProMetIleGluTyrAlaLeuGlyGlyPheGlnProSerGlyProLysGly 421  
Db 1436 CAGAACAGCAGATGATGAATGGGCCCTCGGTGGCTTCAGGCCCTCGGTCTCTAAGGGC 1495  
QY 422 LeuGluProProGluGluLysAsnProTyrLysGluValTyrThrAspMetTyrVal 441  
Db 1496 CTGAACCCACAGAGAGAGAGAAATCCTTACAGGAAGTTTACACCCGACATGTGGGTG 1555  
QY 442 GluProGluAlaAlaAlaTyrAlaProProProAlaLysLysProArgLysSerThr 461  
Db 1556 GAGCTGAAGCAGCTGCTTACGCCCCACCCACAGCAGCAAGAAACCCAGAAAGAGCACA 1615  
QY 462 AlaGluLysProLysValLysGluIleAspGluArgThrArgGluArgLeuValTyr 481

Db 1616 ACAGAGAAAACCTTAAGGTTCAAGGAGATCATTTGATGAGCGCACAAAGGAGCGGCTGTGTAT 1675  
QY 482 GluValArgGlnLysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsn 501  
Db 1676 GAGGTCCGCGAGAGTGCAGAAACATCGAGGACATTTGTATCTCATGTGGGAGCCTCAAT 1735  
QY 502 ValThrLeuGluHisProLeuPheValGlyGlyMetCysGlnAsnCysLysAsnCysPhe 521  
Db 1736 GTACCCCTGGAGCACCACTCTTCAATGGAGGCACTGTGCCAGAACTGTAAAGAACTGCTTC 1795  
QY 522 LeuGluCysAlaTyrGlnTyrAspAspGlyTyrGlnSerTyrCysThrIleCysCys 541  
Db 1796 TTGGAGTGTCTTACAGTATGACGATGGGTACCACTCTATTGACCATCTGCTCT 1855  
QY 542 GlyGlyArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCysPheCysValGlu 561  
Db 1856 GGGGGCGGTGAAGTGTCTCATGTGGGAACAACAACCTGCTCAGGTGCTTTTGTGTGAG 1915  
QY 562 CysValAspLeuValGlyProGlyAlaAlaGlnAlaAlaIleLysGluAspProTyr 581  
Db 1916 TGTGTGGATCTCTTGGTGGGCCAGGAGCTGCTCAGCGAGCAATTAAGGAAGACCCCTGG 1975  
QY 582 AsnCysTyrMetCysGlyHisLysGlyThrTyrGlyLeuLeuArgArgGluAspTyr 601  
Db 1976 AACTGCTACATGTGGGCAATAAGGGCACCTATGGGCTGCTGCCAAGACGGAGACTGG 2035  
QY 602 ProSerArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProLys 621  
Db 2036 CCTTCTCGACTCCAGATGTTCTTTGGCAATAAACCATCAGCAAGAAATTTGACCCCCAAG 2095  
QY 622 ValTyrProProValProAlaGluLysArgLysProIleArgValLeuSerLeuPheAsp 641  
Db 2096 GTTACCACCTGTGCCAGCTGAGAAAGAGAGCCATCCCGTGTCTCTCTTTGAT 2155  
QY 642 GlyIleAlaThrGlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIle 661  
Db 2156 GGGATTGCTACAGGGCTCTCTGGTGTGAAAGACCTGGGCATCCAAAGTGGACCGCTACAT 2215  
QY 662 AlaSerGluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIle 681  
Db 2216 GCCTCCAGGTGTGTCAGGACTCCATCACGTGGGCATGTTGCGGCACAGGAAAGATC 2275  
QY 682 MetTyrValGlyAspValArgSerValThrGlnLysHisIleGlnGluTyrGlyProPhe 701  
Db 2276 ATGTAGCTCGGGACCTCCGACGCTCACAGAAAGCATATCCAGGAGTGGGGCCCATTC 2335  
QY 702 AspLeuValIleGlySerProCysAsnAspLeuSerIleValAsnProAlaArgLys 721  
Db 2336 GACCTGGTGAATGGAGCAGTCCCTGCAATGACCTCTCCATTGTCAACCCCTGCCCGAAG 2395  
QY 722 GlyLeuTyrGluGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAla 741  
Db 2396 GGACTTTATGAGGTHACTGGCCGCTCTTTTGAGTTCTACCGCTCTCTGATGATGCG 2455  
QY 742 ArgProLysGluGlyAspAspArgProPheThrLeuPheGluAsnValValAlaMet 761  
Db 2456 CGGCCCAAGAGGAGATGATCGCCCTCTTCTTGGCTCTTTGAGAAATGTTGGTGGCCATG 2515  
QY 762 GlyValSerAspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAsp 781  
Db 2516 GCGTTAGTGACAGAGGACATCTCGGATTTCTTGAGTCTAACCCCGTGTGATGATGAC 2575  
QY 782 AlaLysGluValSerAlaAlaHisArgAlaArgTyrPheTyrGlyAsnLeuProGlyMet 801  
Db 2576 GCCAAAGAAAGTGTCTGCTGCACACAGGGCCCTTACTTCTGGGGTAACCTTCTCTGGCATG 2635  
QY 802 AsnArgProLeuAlaSerThrValAsnAspLysLeuGluLeuGlnGluCysLeuGluHis 821  
Db 2636 AACAGCCCTTTGGCATCCACTGTGAATGATGAAGTGGAGCTGCAAGAGTGTCTGGAGCAC 2695  
QY 822 GlyArgIleAlaLysPheSerLysValArgThrIleThrArgSerAsnSerIleLys 841  
Db 2696 GCGAAATAGCCCAAGTTACAGAAAGTGAGGACCATTAACCAACAGGTCAAACTCTATAAG 2755

QY 842 GlnGlyLysAspGlnHisPheProValPheMetAsnGluGlyAspIleLeuTrpCys 861  
 DB 2756 CAGGGCAAGACCAAGCATTTCCCGCTTCATGAAACAGAGAGACATCCCTGTGGC 2815  
 QY 862 ThrGluMetGluArgValPheGlyPheProValHisIleThrAspValSerAsnMetSer 881  
 DB 2816 ACTAATAAGGAAGGGGTGGCTCCCGCTCCACATACACAGACGTCTCCCAATGAGC 2875  
 QY 882 ArgLeuAlaArgGlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeu 901  
 DB 2876 CGCTTGGGAGAGAGAGACTGCTGGGCCATGCTGAGCGCGCTCATCCCGCACCTC 2935  
 QY 902 PheAlaProLeuLysGluIleThrPheAlaCysVal 912  
 DB 2936 TTCGCTCCGCTGAAGAAATATTTGCTGTGTG 2968  
 REF 5  
 ACT 24  
 LOCUS AC120824 123936 bp DNA linear HTG 23-JUL-2002  
 DEFINITION Rattus norvegicus clone CH230-282D16, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\* 38 unordered pieces.  
 AC120824.2 GI:21902541  
 HTG: HTGS PHASE1.  
 KEYWORDS Rattus norvegicus.  
 SOURCE Rattus norvegicus.  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 1 (bases 1 to 123936)  
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaralunga,H.C., Are,J.R., Ayele,M., Banks,T., Barabara,J., Benton,J., Bimagne,K., Blankenburg,K., Bonni,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dachorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabris,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Correll,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homsl,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratochvic,J., Kurehni,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W., Louised,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Maxhiney,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,S., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,N., Okwomu,G., Oragunye,N., Oyiedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Fu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., RojuboKan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Umanai,K., Vaequez,L., Vera,V., Villalob,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstock,G., and Gibbs,R.  
 TITLE Direct Submission  
 JOURNAL Unpublished

REFERENCE 2 (bases 1 to 123936)  
 AUTHORS Morley,K.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 123936)  
 Morley,K.C.  
 REFERENCE Direct Submission  
 TITLE Submitted (23-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
 On Jul 18, 2002 this sequence version replaced gi:20531783.  
 COMMENT  
 ----- Genome Center of Medicine  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GMLV  
 Center clone name: CH230-282D16  
 ----- Summary Statistics  
 Sequencing vector: Plasmid  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 95907 bases at least Q40  
 Consensus quality: 99223 bases at least Q30  
 Consensus quality: 101036 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 38 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1  
 1009: contig of 1009 bp in length  
 1010 1109: gap of unknown length  
 1110 2113: contig of 1004 bp in length  
 2114 2213: gap of unknown length  
 2214 3233: contig of 1020 bp in length  
 3234 3333: gap of unknown length  
 3334 4442: contig of 1109 bp in length  
 4443 4542: gap of unknown length  
 4543 6202: contig of 1660 bp in length  
 6203 7464: gap of unknown length  
 7465 7565: gap of unknown length  
 7566 8772: contig of 1208 bp in length  
 8773 8872: gap of unknown length  
 8873 10371: contig of 1499 bp in length  
 10372 10471: gap of unknown length  
 10472 11714: contig of 1243 bp in length  
 11715 11814: gap of unknown length  
 11815 12924: contig of 1110 bp in length  
 12925 13024: gap of unknown length  
 13025 14347: contig of 1323 bp in length  
 14348 14447: gap of unknown length  
 14448 15652: contig of 1205 bp in length  
 15653 15752: gap of unknown length  
 15753 17796: contig of 2044 bp in length  
 17797 17896: gap of unknown length  
 17897 18937: contig of 1041 bp in length  
 18938 19037: gap of unknown length  
 19038 21543: contig of 2506 bp in length  
 21544 21643: gap of unknown length  
 21644 24540: contig of 2897 bp in length  
 24541 28112: gap of unknown length  
 28113 28212: gap of unknown length



QY 646 G1yleuLeuValleuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSerGluVal 665  
Db 86062 GGGCTCCGTGGTGTGAAGGACCTGGGACATCAAGTGAAGCCCTACATTCCTCCGAGGTG 86121  
QY 666 CysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyrValGly 685  
Db 86122 TGCAGAGACTCCATCAACCTGGGCAATGACCGGACCAAGGAAATCATGTCGTGGG 86181  
QY 686 AspValAspSerValThrGlnLysHisIleGlnGluTyrPheProPheAspLeuValIle 705  
Db 86182 GACCTCTGCAGATCAACACAAAGATATCCAGAGAGGGGCCCATACATCTGGTGAATT 86241  
QY 706 GlyGlySerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeuTyrGlu 725  
Db 86242 GGGGCGAGCTCTTCATGACCTCTCCATCATCAACCTCCCGGACGAGGACTTACAG 86301  
QY 726 GlyThrGlyArgLeuPheGluPheTyrArgLeuLeuHisAspAlaArgProLysGlu 745  
Db 86302 GGCACCTGGAGCCTCTTCTTGTAGTCTACACCTCTCTGATGATGCGGCGCCAGAG 86361  
QY 746 GlyAspAspArgProPhePheThrLeuPheGlu 756  
Db 86362 GGAAGAGATCGCCCTTCTGTGTGCTTTGAG 86394  
RESULT 6  
LOCUS AK025230 2191 bp mRNA linear PRI 29-SEP-2000  
DEFINITION Homo sapiens cDNA: FLJ21577 fis, clone COL06724, highly similar to  
AF067972 Homo sapiens DNA cytosine methyltransferase 3 alpha  
(DNMT3A) mRNA.  
ACCESSION AK025230  
VERSION AK025230.1 GI:10437699  
KEYWORDS oligo cloning; fis (full insert sequence).  
SOURCE Homo sapiens colon cDNA to mRNA, clone\_lib:COL clone:COL06724.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Kawabata,A., Hiki,T., Kobatake,N., Inagaki,H., Ikema,Y.,  
Okamoto,S., Okitani,R., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T.,  
Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.  
NEBO human cDNA sequencing project  
TITLE Unpublished  
JOURNAL 2 (bases 1 to 2191)  
REFERENCE 1 (bases 1 to 2191)  
Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,  
Shibahara,T., Tanaka,T. and Nakamura,Y.  
TITLE Direct Submission  
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure Analysis, Human  
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,  
Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
Fax:81-3-5449-5416)  
COMMENT NEBO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan: cDNA full insert  
sequencing: Research Association for Biotechnology: cDNA library  
construction, 5'- & 3'-end one pass sequencing: Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).  
FEATURES  
source location/Qualifiers  
1..2191  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="COL06724"  
/issue\_type="Colon"  
/clone\_lib="COL"  
/note="cloning vector pMB18SFL3"  
misc\_feature 1..2191  
/note="highly similar to AF067972 Homo sapiens DNA  
cytosine methyltransferase 3 alpha (DNMT3A) mRNA"  
BASE COUNT 571 a 505 c 575 g 540 t  
ORIGIN

Alignment Scores:  
Pred. No.: 3.2e-88 Length: 2191  
Score: 2413.00 Matches: 441  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.77% Mismatches: 0  
Query Match: 48.90% Indels: 0  
Gaps: 0  
US-09-720-086-7 (1-912) x AK025230 (1-2191)  
QY 471 IleAspGluArgThrArgGluArgLeuValTyrGluValArgGlnLysCysArgAsnIle 490  
Db 2 ATTATGAGCGGACAGAGAGAGCGGCTGGTGTACAGGTGGCGAGAAATGCCGAACATT 61  
QY 491 GluAspIleCysIleSerCysGlySerLeuAsnValThrLeuGlnHisPheLeuPheVal 510  
Db 62 GAGACATCTGCATCTCTTGGAGAGCTCAATGTACCTGGAACACCCCTCTTCTGTT 121  
QY 511 GlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGluCysAlaTyrGlnTyrAspAsp 530  
Db 122 GAGAGAAATGTGCCAAACTGCCAAGACTGCTTTCTGGAGTGTGGTACAGACGAC 181  
QY 531 AspGlyTyrGlnSerTyrCysThrIleCysCysGlyGlyArgGluValLeuMetCysGly 550  
Db 182 GAGGCTACCACTCCACTCAACCATCTGCTGAGGGCCGTGAGTGTCTCATGTGGCA 241  
QY 551 AsnAsnAsnCysCysArgCysPheCysValGluCysValAspLeuValGlyProGly 570  
Db 242 AACACAACTGCTGAGGTCTTTGGCTGAGTGTGAGACCTTGTGGGGCGGG 301  
QY 571 AlAlaGlnAlaAlaIleLysGluAspProTyrPasnCysTyrMetCysGlyHisLysGly 590  
Db 302 GCTCCCGAGCGAGCCATTAAAGAGACCCCTGGAATCTGCATCATGTGGCGGCAAGGT 361  
QY 591 ThrTyrGlyLeuLeuArgArgGluAspTyrProSerArgLeuGlnMetPhePheAla 610  
Db 362 ACCTACGGGCTGCTGGGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 421  
QY 611 AsnAsnHisAspGlnGlnPheAspProProLysValTyrProProValProAlaGluLys 630  
Db 422 AATAACCAACAGACAGAGAAATTGACCTCCAAAGTTTACCACCTGTCCAGGTGAGAG 481  
QY 631 ArgLysProIleArgValLeuSerLeuPheAspGlyIleAlaThrGlyLeuLeuValLeu 650  
Db 482 AGAAGAGCCATCGGGTGTCTCTCTTGTATGAAATCCTACAGGGCTCTCGTGTCTG 541  
QY 651 LysAspLeuGlyIleGlnValAspArgTyrIleAlaSerGluValCysGluAspSerIle 670  
Db 542 AAGGACTGGGCAATTCAGGTGAGACCGCTACATTGCTCGAGAGTGTGAGGACTCCATC 601  
QY 671 ThrValGlyMetValArgHisGlnGlyLysIleMetTyrValGlyAspValArgSerVal 690  
Db 602 ACGGTGGCATGATGGGCGACCAAGGAAAGATCATGTACGCGGGAGCGTCCGACGGTCT 661  
QY 691 ThrGlnLysHisIleGlnGluTyrPheProPheAspLeuValIleGlyGlySerProCys 710  
Db 662 ACACAGAAAGCATATCCAGAGAGTGGGCGCATTCGATCGATTGGGGCGGAGTCCCTGC 721  
QY 711 AsnAspLeuSerIleValAsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeu 730  
Db 722 AATGACCTCTCCATCGTCAACCTCTGCTGCAAGGGGCTTACGAGGGGCACTGCGCGCTC 781  
QY 731 PhePheGluPheTyrArgLeuLeuHisAspAlaArgProLysGluGlyAspAspArgPro 750  
Db 782 TTCTTTGAGTTTACCGGCTCTCTCATATGCGGCGCCAAAGAGGAGATGATCGCCCC 841  
QY 751 PhePheThrPhePheGluAsnValValAlaMetGlyValSerAspLysArgAspIleSer 770  
Db 842 TTCTTCTGCTCTTTAGAAATGTGTGCGCATGGGCGCTTGTGTGACGAGGAGCATCTCG 901  
QY 771 ArgPheLeuGluSerAsnProValMetIleAspAlaLysGluValSerAlaHisArg 790







Db 768 GCTACGAGCTCTGGAGACGCTCA----- 791  
Qy 216 LysLysAlaLysValIleAlaGlyMetAsnAlaValGluLysAsnGlnGlyProGlyGlu 235  
Db 792 -----GCATCGCTCTCAAGACGACGCGCATGGCATCCCTGCCAGC 833  
Qy 236 SerGlnLysValGluGlnLysSerProProAlaValGlnGlnProThrAspProAlaSer 255  
Db 834 GTCGACTTCATGAGAGAGTACACCTTAGAGGCTCACTACCCA-----TCA 881  
Qy 256 ProThrValAlaThrThrProGluProValGlySerAspAlaGlyAspLysAsnAlaThr 275  
Db 882 GTTACTTGCAGCAGATGAGATCAGAGGGATATGATACACACAGGATGACAG 941  
Qy 276 LysAlaGlyAspAspGluProGluTyrGluAspGlyArgGlyPheGlyIleGlyGluLeu 295  
Db 942 AGCATATATGAGACACACAGAGATACAGATGATGAAGAGTTGGAAATAGTGACTC 1001  
Qy 296 ValTyrGlyLysLeuArgGlyPheSerTyrTyrProGlyArgGlyIleValSerTyrTyrMet 315  
Db 1002 GTGTGGGAAAGATCAAGGGCTTCTCTGCTGGCTGCATGGTGTCTTGGAAAGCC 1061  
Qy 316 ThrGlyArgSerArgAlaAlaGluGlyThrArgTyrValMetTyrPheGlyAspGlyLys 335  
Db 1062 ACCCTCAAGCAGACAGCCATGCCGGAATGCCGTGGTACAGTGGTTGGTGTGAGGCAAG 1121  
Qy 336 PheSerValValCysValGluLysLeuMetProLeuSerSerPheCysSerAlaPheHis 355  
Db 1122 TTTTCTGAGATCTCTGTCGACAACTGGTGGCTCTGGGGCTGTTCAGCCAGCACTTAT 1181  
Qy 356 GlnAlaThrTyrAsnLysGlnProMetTyrArgLysAlaIleTyrGluValLeuGlnVal 375  
Db 1182 CTGGCTACCTTCATTAAGCTGTTCTTATAGGAAGCCATGTATACCACTCTGAGAAA 1241  
Qy 376 AlaSerSerArgValGlyLysLeuPheProValCysHisAspSerAspGluSerAspThr 395  
Db 1242 GCCAGGGTTCGAGCTGCACAAAGCTTC-----TCCAGCAGTCT 1280  
Qy 396 AlaLysAlaValGluValGlnAsnLysProMetIleGluTyrAlaLeuGlyGlyPheGln 415  
Db 1281 GGAGAGTCACTGGAGGACCACTGAAGCCATGCTGAGTGGGCCACGGTGGCTTCAAG 1340  
Qy 416 ProSerGlyProLysGlyLeuGluProProGluGluLysLysAsnProTyrLysGlu 434  
Db 1341 CCTACTGGATCAGGGCCTCAACCCAAAGAACGACCGAGAACAAAGTCAAGA 1400  
Qy 435 ValTyrThrAspMetTyrValGluProGluAlaAlaIleTyrAlaProProProAla 454  
Db 1401 CGACCAACCAAT-----GACTCTGCTCTTGAATCCCCCAACCCCAAG 1445  
Qy 455 LysLysProArgLysSerThrAlaGluLysProLysValLysGluIleIleAspGluArg 474  
Db 1446 CGCCTCAAGACAATATGCTATGCGGGAAGAGCCGAGGAG-----GATGAGAG 1496  
Qy 475 ThrArgGluArgLeuValTyrGluValArgGlnLysCysArgAsnIleGluAspIleCys 494  
Db 1497 AGCCGAGAACGAGTGGCTTCTGAGTCAACCAACAAAGGCAATCTGGAAGACCGCTGT 1556  
Qy 495 IleSerCysGlySerLeuAsnValThrLeuGluHisProLeuPheValGlyGlyMetCys 514  
Db 1557 TTGTCTTGGAAGAAAGAACCTGTGTCTTCAACCCCTCTTGAAGGGTGGCTGTGT 1616  
Qy 515 GlnAsnCysLysAsnCysPheLeuGluCysAlaTyrGlnTyrAspAspArgLysTyrGln 534  
Db 1617 CAGAGTTGCGGGATCGCTTCTTAGAGCTTCTTACATGATGATGAGGACGGCTATAG 1676  
Qy 535 SerTyrCysThrIleCysCysGlyArgGluValLeuMetCysGlyLysAsnAsnCys 554  
Db 1677 TCTTACTGACCGGTGTGTGTAGGCGCGGAACGTGCTGTGAGTAAACACAGCTGC 1736  
Qy 555 CysArgCysPheCysValGluCysValAspLeuValGlyProGlyAlaAlaGlnAla 574

Db 1737 TGCAGATGCTTCTGTGTGAGATGTCTGAGAGTGTCTGGGCGGACAGCAGCTGAGAT 1796  
Qy 575 AlaIleLysGluAspProTyrPasnCysTyrMetCysGlyHisLysGlyThrTyrGlyLeu 594  
Db 1797 GCCAAGCTGCAGAACCTTGGAGCTGTATATGTGCTCTCCCTCAAGCGCTGCATGGGCTC 1856  
Qy 595 LeuArgArgArgGluAspTyrProSerArgLeuGlnMetPhePheAlaAsnAsnHisAsp 614  
Db 1857 CTCGACGAGAGAAAGTTGGAAACATGCGCCCTGCAAGACTTCTTCACTCATGTATCTGAC 1916  
Qy 615 ---GlnGluPheAspProProLysValTyrProProValProAlaGluLysArgLysPro 633  
Db 1917 CTGAAAGATTTGAGCACCAAGTTGTATCCAGCAATTCCTGACGCAAAAGAGAGGCC 1976  
Qy 634 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyLeuLeuValLeuLysAspLeu 653  
Db 1977 ATTAGATCTCTGTCTGTGTGATGGAATTGCAACGGGGTACTTGTGCTCAAGAGTTG 2036  
Qy 654 GlyIleGlnValAspArgTyrIleAlaSerGluValCysGluAspSerIleThrValGly 673  
Db 2037 GGATTTAAAGTGAAGAAAGTACATTGCTCCGAAAGTCTGTGACAGATCCATCGCTGGGA 2096  
Qy 674 MetValArgHisGlnGlyLysIleMetTyrValGlyAspValArgSerValThrGlnLys 693  
Db 2097 ACTGTTAAGCATGAAGCCAGATCAATATGTCAATGACGTCGGAAATACCAAGAAA 2156  
Qy 694 HisIleGlnGluTyrGlyProPheAspLeuValIleGlyGlySerProCysAsnAspLeu 713  
Db 2157 AATATTGAAGAGTGGGCGCCGCTGCACTTGATGTGTGGAAGCCCATGCAATGATCTC 2216  
Qy 714 SerIleValAspProAlaArgGlyGlyLeuTyrGluGlyThrGlyArgLeuPheGlu 733  
Db 2217 TCTAACGTCAATCTCTGCCGCAAGGTTATATGAGGCAACAGGAAGCTCTTCTTCCAG 2276  
Qy 734 PheTyrArgLeuLeuHisAspAlaArgProLysGluGlyAspAspArgProPhePheTyr 753  
Db 2277 TTTTACCACTTGTGAAATTATACCCGCCCAAGAGGCGCAACCGTCAATCTTCTCGG 2336  
Qy 754 LeuPheGluAsnValValAlaMetClyValSerAspLysArgAspIleSerArgPheLeu 773  
Db 2337 ATGTTGAGAACTTGTGGCCATGAAAGTGAATGACAAAGAAAGATCTCAAGATCTCTG 2396  
Qy 774 GluSerAsnProValMetIleAspAlaLysGluValSerAlaAlaHisArgAlaArgTyr 793  
Db 2397 GCATGTAAACCCAGTGAATGATGATGCCATCAAGGTGTCTGCTGCACAGGGCCCGTAC 2456  
Qy 794 PheTyrGlyAsnLeuProGlyMetCAsnArgProLeuAlaSerThrValAsnAspLysLeu 813  
Db 2457 TTCTGGGGTAACTTACCCGGATGAACAGGCCCGGTATGCTTCAAGAAATGATTAAGCTC 2516  
Qy 814 GluLeuGlnGluCysLeuGluHisGlyArgIleAlaLysPheSerLysValArgThrIle 833  
Db 2517 GAGCTGCAGAGACTCGCTGAGTTCAGTAGACAGCAAGTTAAAGAAATGACACATA 2576  
Qy 834 ThrThrArgSerThrAsnIleLysGlnGlyLysAspGlnHisPheProValPheMetAsn 853  
Db 2577 ACCACCAAGTGCAACTTCATCAGACAGGCAAAACAGACTTTTCTCTGTAGTCATGAT 2636  
Qy 854 GluLysGluAspIleLeuTyrProCysThrGluMetGluArgValPheGlyPheProValHis 873  
Db 2637 GGCAGAGAGAGACTTTGTGTGCTGAGCTGAAAGGATCTTGGCTTCCCTGCTCAC 2696  
Qy 874 TyrThrAspValSerAsnMetSerArgLeuAlaArgGlnArgLeuGluArgSerTyr 893  
Db 2697 TACACGAGAGTGTCCAACTATGGCGCGCGCGCTGTACAGACTGTGGCAGGTCTCGG 2756  
Qy 894 SerValProValIleArgHisLeuPheAlaProLeuLysGluTyrPheAlaCys 911  
Db 2757 AGGTACCGGTATCACAACACTGTTTGGCCCTTGAAGGACTACTTGTCTGT 2810  
RESULT 8  
AC009474  
LOCUS AC009474 168651 bp DNA linear PRI 08-NOV-2000



```
repeat_region 26074..26366 /rpt_family="Alu"
repeat_region 26384..26542 /rpt_family="MIR"
repeat_region 26569..27058 /rpt_family="L2"
repeat_region 27140..27647 /rpt_family="L2"
repeat_region 30824..31150 /rpt_family="MALR"
repeat_region 32262..32525 /rpt_family="CR1"
repeat_region 32568..32770 /rpt_family="MIR"
repeat_region 32775..33032 /rpt_family="MIR"
repeat_region 33172..33418 /rpt_family="MIR"
repeat_region 34930..35230 /rpt_family="Alu"
repeat_region 35302..35555 /rpt_family="L2"
repeat_region 35975..36081 /rpt_family="L2"
repeat_region 36364..36515 /rpt_family="Alu"
repeat_region 37266..37340 /rpt_family="L2"
repeat_region 37400..37440 /rpt_family="L2"
repeat_region 37615..37715 /rpt_family="MIR"
repeat_region 38006..38177 /rpt_family="MIR"
repeat_region 43171..43204 /rpt_family="MIR"
repeat_region 44777..44926 /rpt_family="MIR"
repeat_region 45905..46110 /rpt_family="MIR"
repeat_region 48479..48638 /rpt_family="MIR"
repeat_region 48914..49033 /rpt_family="MIR"
repeat_region 49069..49337 /rpt_family="MIR"
repeat_region 49340..49621 /rpt_family="Alu"
repeat_region 50377..50498 /rpt_family="MIR"
repeat_region 56393..56545 /rpt_family="MIR"
repeat_region 58109..58420 /rpt_family="MIR"
```

Alignment Scores:

Pred. No.:	4.54e-79	Length:	168651
Score:	2227.50	Matches:	488
Percent Similarity:	68.35%	Conservative:	67
Best Local Similarity:	60.10%	Mismatches:	137
Query Match:	45.14%	Indels:	131
DB:	9	Gaps:	9

US-09-720-086-7 (1-912) x AC009474 (1-168651)

```
QY 210 Trrlysaarglualaglyslvalalysvalilialaglymetasnalavalglu 229
Db 74844 TGGCTCATGGCTGAGAGAGCGCGAGTCACCTGATGATGATGATGATGAGAA 74903
QY 230 AsnGlnGlyProGlyGlnSerGlnValGlnGlnGlnGlnGlnGlnGlnGln 249
Db 74904 AACTAGAGCTCAGAGAGGTGCAAAAGGTGAGAGAGCCAGCCCTGCTGCGGAG 74963
QY 250 ProthrAapProAlaSerProThrValAlaThrThrProGluProValGlySerAspAla 269
```

```
Db 74964 CCCAGCAACCCCATCCCTCCCTGTTACCAACACCTGAGCTGTGGGAGCTGAGGCT 75023
QY 270 GlyAspLysAsnAlaThrLysAlaGlyAspAspGluProGluTyGlnuAspGlyArgGly 289
Db 75024 GGGAAACAAG--TACATCAAAAGCAGATGATGATGATGATGATGATGATGATGATG 75080
QY 290 PheGlyIleGlyGlnuValTrpGlyLysLysLysLysLysLysLysLysLysLys 309
Db 75081 TTTGGCAATGGGAGCTGCTGTGAGGAAATCTGAGGCTTCTCTGCTGCTGCTGCTGCT 75135
QY 310 IleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGlnGlyThrArgTrpValMet 329
Db 75136 ATTGTGCTTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 75195
QY 330 TrpPheGlyAspGlyLysPheSerValValCysValGlnLysLysMetProLysSer 349
Db 75196 TGGCTTGAGA--GCTGAGTCTCAGTCTGCTGTGAGAGAGCTGATACCTGAGCTCC 75254
QY 350 PheCysSerAlaPheIleGlnAlaThr-----TyrAsnLysGlnProMetTyrArgLys 367
Db 75255 TTTGCAAGTGTGTTCCACCAAGCCAC--CTATAATACCAAGAGCCCATGTACCAAAA 75313
QY 368 AlaIleTyGlnValLeuGlnValAlaLysSerArgAlaGlyLysLeuPheProValCys 387
Db 75314 GCCATCTAGAACTCTGCGCGGTGCGCAGCAGTGT--ACAGGAGAGCTGTTCTCAGGAGC 75372
QY 388 HisAspSerAspGluSerAspThrAlaLysAlaValGlnValGlnAsnLysProMetIle 407
Db 75373 CATGACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 75432
QY 408 GluTrpAlaLeuGlyGlyPheGlnProSerGlyProLysGlyLeuGluProProGlnGlu 427
Db 75433 AGATGAGACTTGTGGGGGTTCCAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 75492
QY 428 GluLysAsnProTrpLysGlnValTyThrAspMetTrpValGluProGluAlaAla 447
Db 75493 GAGAAGATCCCAAGAAAGATTATGACACGTGAGGGGTGACCTGATGAGCACTGCT 75552
QY 448 TyrAlaProProProProAlaLysLysProArgLysSerThrAlaGlnLys----- 464
Db 75553 TAGGACACCAACCCCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 75612
QY 464 ----- 464
Db 75613 GACACGACAGTAGCTACACCTGATATCCAGACCTTGGAGGCTGAGGCGGGCGG 75672
QY 464 ----- 464
Db 75673 ATCAGAGCTCAGAGATAGACCTGACATGTCAGGCCAACAATGTGAAACCCGCT 75732
QY 465 ----- 465
Db 75733 CTACTAAAAAGTACCAAAATAGCTGACATGATGATGATGATGATGATGATGATG 75792
QY 466 ----- 466
Db 75793 GAGAGGCTGAGGAGAGATCACTTGAACAGAGAGTCGAGAGTTACAGTGAAGCGAGA 75852
QY 466 ----- 466
Db 75853 TAGCACCATTCACCTCCAGCTGTGAGACAGCAAGACTCTGTTCAAAAAATAAAA 75912
QY 467 ----- 467
Db 75913 AAAAAAGCTGACATCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 75970
QY 482 uValArgGlnLysCysArgAsnIleGlnAspIleCysIleSerCysGlySerLeuAsnVa 502
Db 75971 AGTGAAGAAAGATGCGAACAATCAAGACCTTGTGTCTCTGTGAGAGCTCAATGT 76030
QY 502 lThrLeuGlnHisProLeuPheValGlyGlyMetCysGlnAsnLysLysAsnLysPheLe 522
```

Qy 881 rArgLeuAlaArgGlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLe 901  
Db 77138 CCATTGGCAGGCGAGACTCTTGGCTAGTTCATGGAGCATAGCAGTCACTTGCATTT 77197  
Qy 901 uPheAlaProLeuLysGluTyPheAlaCysVal 912  
Db 77198 CTTCCCTCCACTGAAGGGGTATTTCTGTTGTG 77231

RESULT 9  
AF068627  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
PubMed  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REMARK  
COMMENT  
FEATURES  
source  
gene  
misc\_feature  
CDS

AF068627 4135 bp mRNA linear ROD 06-DEC-1999  
Mus musculus DNA cytosine-5 methyltransferase 3B2 (Dnmt3b) mRNA,  
alternatively spliced, complete cds.  
AF068627  
AF068627.2 GI:6449471  
Mus musculus  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1 (bases 1 to 4135)  
Okano,M., Xie,S. and Li,E.  
Cloning and characterization of a family of novel mammalian DNA  
(cytosine-5) methyltransferases  
Nat. Genet. 19 (3), 219-220 (1998)  
96324766  
9662389  
2 (bases 1 to 4135)  
Xie,S., Okano,M. and Li,E.  
Direct Submission  
Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,  
Charlestown, MA 02129, USA  
3 (bases 1 to 4135)  
Okano,M., Chijiwa,T., Sasaki,H. and Li,E.  
Direct Submission  
Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,  
Charlestown, MA 02129, USA  
Sequence update by submitter  
On Nov 18, 1999 this sequence version replaced gi:3327981.  
Location/Qualifiers  
1..4135  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/chromosome="2"  
/map="84.0 cM"  
1..4135  
/gene="Dnmt3b"  
/note="Dnmt3b"  
/note="Dnmt3b"  
/note="similar to EST sequences deposited in GenBank  
Accession Numbers AA116694, AA119979, AA177277, AA210568,  
AA407106, and AA575617"  
269..2788  
/gene="Dnmt3b"  
/function="de novo DNA methylation"  
/note="alternatively spliced product; contains Cys-rich  
region; C-terminal region is similar to corresponding  
region of DNA cytosine-5 methyltransferase 3A"  
/codon\_start=1  
/product="DNA cytosine-5 methyltransferase 3B2"  
/protein\_id="AAC40179.2"  
/db\_xref="GI:6449472"  
/translation="MKGDSRHLNEEGASQVEECII VNGNFSDOSSDTKADPSPVLE  
AICPEVCTPTEGRSSRLSKREYSLILNYTQDMTGDGRDDEVDGNGSDILMPK  
LRTKQDPTRESAPVTRFHSNGTSLSLEQRASPRITRGQGRHHVQEPVPEFPATR  
SRRRASSASTPNSPASVDMEVTPKSVTSFVSLQSDQEGMTTQVDAEGRD  
GDSYTDQDFEGIGLGVKGLKGFNSWPAVSWTKARQKPMARVQWFGDKP  
SEISADKLWALGFLGSFNLATFKLVSYAKAMTILEKARVAKCTKTSFSSGEEGL  
QLKPMLEWAHGFSKEVTIGELNKKPKPENKRRRTNDSAAESPGLKRLTNSYGL  
KORGEDESRRMASEVTINKNLEDRICSGKNKPVSFHPLFEGELGQADKLQEPWS  
FYMDEDQYQYCTVCCGECRELLSCRCFCVECLVLVGAETADKLQEPWS  
CYMCLPQRCGVLRNRKNWMLQDFTTDPDLEEFPPKLYPAIPAAKRRRTIRVLISL



```
QY 575 AlalLeuGluAspProTrpAsnCyseTyrMetCysGlyHisLysGlyThrTyrGlyLeu 594
Db 1769 GCCAAGCTGAGAACCTGGAGCTCTATATGTTGCTCCCTCAGCGCTGCTGGGTC 1828
QY 595 LeuArgArgGluAspTrpProSerArgLeuGluMetPhePheAlaAsnAsnHisAsp 614
Db 1829 CTCGACGAGAAAGATTGGAACATGCGCTGCAAGACTTCTTCACTACTGCTGAC 1888
QY 615 ---GlnGluPheAspProProLysValTyrProProValProAlaGluLysArgLysPro 633
Db 1889 CTGGAGAAATTGAGCCACCAAGTTGTACCCAGCAATCTCTGCGAGCAAAAGAGGCC 1948
QY 634 IleArgValLeuSerLeuPheAspGlyIleAlaThrGlyLeuLeuValLeuLysAspLeu 653
Db 1949 ATTAGAGTCTGCTCTGTTGATGGAATTGCAACGGGGTACTTGGTGCTCAAGAGATTG 2008
QY 654 GlyIleGlnValAspArgTyrIleAlaSerGluValCysGluAspSerIleThrValGly 673
Db 2009 GGTATTAAAGTGAAGAGTACATTGCTCCGAAGTCTGTGCAGAGTCCATCGCTGGGA 2068
QY 674 MetValArgHisGlnGlyLysIleMetTyrValGlyAspValArgSerValThrGlnLys 693
Db 2069 ACTGTTAAGCATGAAGGCCAGATCAATATGTCAATGACGTCCGGAATAATCACCAAGAA 2128
QY 694 HisIleGlnGluTrpGlyProPheAspLeuValIleGlyGlySerProCysAsnAspLeu 713
Db 2129 AATATTGAAGAGTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2188
QY 714 SerIleValAsnProAlaArgLysGlyLeuTyrGluGlyThrGlyArgLeuPheGlu 733
Db 2189 TCTAACGTCATCTCTGCGCAAGGTTTATATGAGGCGACAGAGGCTCTTCTTCTGAG 2248
QY 734 PheTyrArgLeuLeuHisAspAlaArgProLysGluGlyAspArgProPheTrp 753
Db 2249 TTTTACCACCTGCTCAATATATACCGCGCCAGAGGCGGACAAACCGTCCATCTCTG 2308
QY 754 LeuPheGluAsnValAlaMetGlyValSerAspLysArgAspIleSerArgPheLeu 773
Db 2309 ATGTTGAGATGTTGTGGCCATGAAGTGAATGACAGAAAGACATCTCAAGATTTCTG 2368
QY 774 GluSerAsnProValMetIleAspAlaLysGluValSerAlaAlaHisArgAlaArgTyr 793
Db 2369 GCATGTAAACCCAGTATGATCATGATCCATCAAGTGTGCTGCTCTACAGGCGCGGTAC 2428
QY 794 PheTrpGlyAsnLeuProGlyMetAsnArgProLeuAlaSerThrValAsnAspLysLeu 813
Db 2429 TTCTGGGGTAACCTACCGGAATGAACAGCGCGGTGATGCTTCAAGAAATGATAGCTC 2488
QY 814 GluLeuGlnGluCysLeuGluHisGlyArgIleAlaLysPheSerLysValArgThrIle 833
Db 2489 GAGCTGCAGGACTGCTGAGTTCAGTAGGACAGCAAGTAAAGAAAGTGCAGACAAATA 2548
QY 834 ThrThrArgSerAsnSerIleLysGlnGlyLysAspGlnHisPheProValPheMetAsn 853
Db 2549 ACCACCAAGTCCAACTCCATCAGACAGGCGCAAAACACGCTTTTCCCTGTAGTCATGA 2608
QY 854 GluLysGluAspIleLeuTrpCysThrGluMetGluArgValPheGlyPheProValHis 873
Db 2609 GGCAGGACGAGCTTTGTGGTGGCTAGCTAGCTGAGAAAGATCTCGGCTTCCCTGCTCAC 2668
QY 874 TyrThrAspValSerAsnMetSerArgLeuAlaArgGlnArgLeuGlyArgSerTrp 893
Db 2669 TACACGACGCTGCTCAACATGGCGCGCGCGCGCTGAGAAAGTGTGGGCGAGTCTCTGG 2728
QY 894 SerValProValIleArgHisLeuPheAlaProLeuLysGluTyrPheAlaCys 911
Db 2729 AGTGACCGGTCATCAGACACCTGTTGCTCCCTTGAAGGACTACTTTGCTGT 2782
RESULT 10
AF151970 4278 bp mRNA linear ROD 08-JUN-2000
LOCUS Mus musculus DNA cytosine-specific methyltransferase isoform 2
DEFINITION (Dnmt3b) mRNA, complete cds.
```

```
ACCESSION AF151970
VERSION AF151970.1 GI:8347119
KEYWORDS Mus musculus.
SOURCE Mus musculus.
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4278)
AUTHORS Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and
Shen, Y.
TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing
isoforms in mouse embryonic tissue
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4278)
AUTHORS Yin, B., Chen, Y. T., Zhu, M., Luo, Y. J., Zhu, N., Xu, S. C., Wu, G. Y. and
Shen, Y.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular
Biology, Institution of Basic Medical Sciences, Chinese Academy of
Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,
P.R.China
FEATURES
source
1..4278
/organism="Mus musculus"
/strain="KM"
/db_xref="taxon:10090"
/dev_stage="8-9 day old embryo"
1..4278
/gene="Dnmt3b"
412..2931
/gene="Dnmt3b"
/function="methylates cytosine in DNA"
/note="alternatively spliced"
/codon_start=1
2..
/product="DNA cytosine-specific methyltransferase isoform
1"
/protein_id="AAF74516.1"
/db_xref="GI:8347120"
/translation="MKGDSRHNLNEEGASGYEELIIVNGFSDOSSDTKADPPSPVLE
AICTFVCTPTGRKRSSRLKREVSSLLNYTQDMTGGDRDDDEVDGNGSDILMPK
LTREFTKTRTRESFAVTRHNSGTSLQRASPRITRQGRHGHVQVEYVEFPATR
SRRRASASATPSPASVPMEEVTPKSVSTPSVDLSQDQDQDQDMDTQVDAESIY
GDSTYQDDKEFGIDGLVMGKIFGFWPMAMVYWKATSKRQAMPQWQVFGDGKF
SEISADLVALGFLSOHFNKLVSYKAMVHTLEKARVAKGTFSSPAGESLED
OLKPMLEWAHGGFKPTGIEGLKPNKKOPENKSRRTTNDAAASPPKRLKTSYC
KRDGEDSRERMASEVTNNKGNLEDRCLSCGKNPVSPHLPFGGLCQSCDRRL
PYMVDGYSQYCTVCCGRELCCSNTSCRCVCYCLEVLVGCAGTAEDAKLQEPWS
CYMCLPQCHGVLRKKDNNMLQDFTTDPDLEEFPPKLYPAIPAKRRPIRVLSL
FDGATGILVLKELGIVKEKYTASEVCASIAVGTVHGGQIKYVNDVRKITKNIEE
WGFPLVIGGSPCNDLSNVNPAKGLYEGTGFVFFYHLNLYTRPKEDNRPFPMF
ENVAMKYNDDKDISRLACNPMDIAIKVSAARARYFWNLFGMRPVMASKNDKL
ELQDLEFSRTAKLKKVQITTKNSIRQGNQLFPVVMNGKDDVLWCTELERIFGFP
AHYTDVSNMGRGAKLGRSVSPVIRHLFADLKDYFACE"
BASE COUNT 1055 a 1113 c 1154 g 956 t
ORIGIN
Alignment Scores:
Pred. No.: 1,83e-80 Length: 4278
Score: 2225.00 Matches: 461
Percent Similarity: 59.24% Conservative: 129
Best Local Similarity: 46.23% Mismatches: 251
Query Match: 45.09% Indels: 155
DB: 10 Gaps: 17
US-09-720-086-7 (1-912) x AF151970 (1-4278)
QY 2 ProAlaMetProSer---SerGlyProGlyAspThrSerSerSer-----AlaAlaGlu 18
Db 147 CCGGACGCGCGGGTAAAGCGCCCAAGTAAACGTAAGCGAGCATCGCGCGCGAGAT 206
QY 19 -ArgGluGluAspArgLysAspGlyGluGluGlnGluProArgLysGluGluArg 38
Db 207 TCGCAACCCGAC-----ACTCCGCG 227
```



OY 38 gGluGluProSerThrThrAlaArgLysValGlyArgProGlyArgLysArgHisPr 58  
DB 228 CCGCCCGCCGCGGAGCCCGCGCGCGATCGCGCGCGCTACAGCAGCTTCACGA 287  
OY 58 o---ProValGluSerGlyAspThrProLysAspProAlaValIle----- 72  
DB 288 CAGGCCCGCTGAGGCTTGCCAGACCTTGAAACCTCAGATAATACCTTCCAGACGC 347  
OY 73 -----SerLysSerProSerMetAlaGlnAs 81  
DB 348 GGGATCTCCCTCCCATCATAGTGCCTTGAGACCAATCCAGGCGCTTCTTTCAGGA 407  
OY 81 p-----SerGlyAlaSerGly 86  
DB 408 AACATGAAGGAGACACGACACATCTGAAATGAAGAGAGGTGCCAGCGGTATGAGA 467  
OY 86 u---LeuLeuProAsnGlyAspLeuGluLysArgSerGluProGlnProGluGluLys 105  
DB 468 GTGCATTATCGTTAAATGGAACTTCAGTACAGTCTCTCAGACACAGAGATGCTCCCTC 527  
OY 105 rProAlaGlyGlyGlnLysGlyAlaProAlaGluGlyGluGlyAlaAlaGluThrIle 125  
DB 528 ACCC----- 531  
OY 125 uProGluAlaSerArgAlaValGluAsnGlyCysCysThrProLys----- 140  
DB 532 -CCAGTCTTGAGGCAATCTGCACAGAGCCAGTCTGCACACCAAGACAGAGCGCGAG 590  
OY 140 ----- 140  
DB 591 GTCAAGCTCCGCGCTGTCTAAGAGGAGGTCTCCAGCTTCTGAATTACAGCGAGACAT 650  
OY 141 -----GluGlyArgGlyAlaProAlaGluAl 149  
DB 651 GACAGAGATGAGACAGAGATGATGAAGTATGATGGAATGGCTCTGATATTCTTAAT 710  
OY 149 agLYLeuGluGlnLysGluThrAsnIleGluSerMetLysMetGluGlySerArgLysAr 169  
DB 711 GCCAAGCTCACCCGTAGACCAAGACACAGAGCGGCTCTGAAAGCCGCGGTCTCG 770  
OY 169 gLeuArgGlyGlyLeuGlyTyrGluSerSerLeuArgLysArgProMetProArgLeuTh 189  
DB 771 AACCGCACTAGCAATGGAGCTTCCAGCTTGAGAGGCAAGAGCTCCCTCCAGAAATCAC 830  
OY 831 r-----PheGlnAlaGlyAspProTyrTyrIle 198  
DB 831 CCGAGGTGCGGAGGCGCGCCACCATGTGCAAGAGTACCCTGTGAGATTCCGGCTACACG 890  
OY 198 eSerLysArgLysArgAspGluTyrLeuAlaArgTyrLysArgGluAlaGluLysAl 218  
DB 891 GTCTCGGAGACGTGCA----- 906  
OY 218 aLysValIleAlaGlyMetAsnAlaValGluGluAsnGlnGlyProGlyGluSerGlnLys 238  
DB 907 -----GATCGCTTACGACAGCAGCCCATGTGATCCCTGCGCAGCGTGCAGCTT 956  
OY 238 sValGluGluAlaSerProProAlaValGlnGlnProThrAspProAlaSerProThVa 258  
DB 957 CATGGAAGAAGTACACCTAGAGGCTCAGTACCCCA-----TCAGTTGACTT 1004  
OY 258 lAlaThrThrProGluProValGlySerAspAlaGlyAspLysAsnAlaThrLysAlaG 278  
DB 1005 GAGCAGAGATGAGATCAGAGGAGTATGATACACACAGGTGATGACAGAGACATATA 1064  
OY 278 yAspAspGluProGluTyrGluAspGlyArgGlyPheGlyTyrIleGlyGluLeuValTyrG 298  
DB 1065 TGGAGACAGACAGAGATCAGAGATGATAAAGAGTTTGGATAGTGACCTCGTGGGG 1124  
OY 298 yLysLeuArgGlyPheSerTyrTyrProGlyArgLysIleValSerTyrTyrMetThrGlyAr 318  
DB 1125 AAAGATCAAGGAGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1184

OY 318 gSerArgAlaAlaGluGlyThrArgTyrValMetTyrPheGlyAspGlyLysPheSerVa 338  
DB 1185 GCGACAGGCCATATGCCGGAATGCCCTGGGTACAGTGGTTGGTGGCAAGTTTCTGA 1244  
OY 338 lValCysValGlnLysLeuMetProLeuSerSerPheCysSerAlaPheIleGlnAlaTh 358  
DB 1245 GATCTCTCTGCAAACTGGTGGCTTGGGCTTTCAGCCAGCATTTAATGGGTAC 1304  
OY 358 rTyrAsnLysGlnProMetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSe 378  
DB 1305 CTTCAATTAAGCTGTTTCTTAATAGGAAGGCCATGTACACACTTTCAGAAAGCCAGGCT 1364  
OY 378 rArgAlaGlyLysLeuPheProValCysHisAspSerAspGluSerAspThrAlaLysAl 398  
DB 1365 TCGAGCTGCGAAGACCTTC-----TCCAGCACTCTCGAGAGCTC 1403  
OY 398 aValGluValGlnAsnLysProMetIleGluTyrAlaLeuGlyGlyPheGlnProSerG 418  
DB 1404 ACTGGAAGACCGCTCAAAAGCCCAACAAAGAACACACAGAACAAAGTCCGAAGCCACAC 1463  
OY 418 yProLysGlyLeuGluProProGluGluGluLysAsnProTyrLysGluValTyrTh 437  
DB 1464 GATGAGAGGCTCTCAAAAGCCCAACAAAGAACACACAGAACAAAGTCCGAAGCCACAC 1523  
OY 437 rAspMetTyrValGluProGluAlaAlaAlaTyrAlaProProProAlaLysLysPr 457  
DB 1524 CAAT-----GACTTGTCTGCTTGTAGTCCCTCCACCAAGCCCTCA 1568  
OY 457 oArgLysSerThrAlaGluLysProLysValLysGluIleLeuAspGluArgThrArgL 477  
DB 1569 GACAAATAGCTATGGCGGGAAGACCGAGGAG-----GATGAGAGAGCCGAGA 1619  
OY 477 uArgLeuValTyrGluValArgGlnLysCysArgAsnIleGluAspIleCysIleSerCy 497  
DB 1620 ACGAGATGCTTCTGAATGATCAACCAACAAAGGCAATCTGGAAGCCGCTGTTTGTCTG 1679  
OY 497 sGlySerLeuAsnValThrLeuGluHisProLeuPheValGlyGlyMetCysGlnAsnCy 517  
DB 1680 TGGAAAGAAAGAACCTGTGTCTTCCACCCCTCTTGAAGGTGGCTGTGTAGACTTG 1739  
OY 517 sLysAsnCyPheLeuGluGluCysAlaTyrGlnTyrAspAspGlyTyrGlnSerTyrCy 537  
DB 1740 CCGGATGCGCTTCTTGAAGCTCTTCTATGATGATGATGAGAGCCGCTATCAGTCTCTG 1799  
OY 537 sThrIleCysCysGlyGlyArgGluValLeuMetCysGlyAsnAsnAsnCysCysArgCy 557  
DB 1800 CACCGTGTCTGTAGAGGCGCGGTGAACTGCTGTGTCAGTAAACAAAGTGTGCAAGATG 1859  
OY 557 sPheCysValGluCysValAspLeuLeuValGlyProGlyAlaAlaGlnAlaIleLys 577  
DB 1860 CTTCTGTGTGAGTGTCTGTGAGGTGTGTGTGGCGGACAGCAGCTGAGATGCCAAGCT 1919  
OY 577 sGluAspProTyrAsnCySerTyrMetCysGlyHisLysGlyThrTyrGlyLeuLeuArgAr 597  
DB 1920 GCAGGAACCTCGAGGTGCTATATGTGCTCTCCCTCAGCGGTGCGATGGGTCTCCGAGC 1979  
OY 597 gArgGluAspTyrProSerArgLeuGluMetPhePheAlaAsnAsnHisAsp---GlnG 616  
DB 1980 CAGGAAGATTTGGAACATGCGCGCTCGAAGACTTCTTCACTACTGATCTCGACTGGAAGA 2039  
OY 616 uPheAspProProLysValTyrProProValProAlaGluLysArgLysProIleArgVa 636  
DB 2040 ATTGAGCCACCCAACTTGTATCCAGCAATTCCTGCAAGCCAAAGAGAGCCCATTAAGT 2099  
OY 636 lLeuSerLeuPheAspGlyIleAlaThrGlyLeuLeuValLeuLysAspLeuGlyIleG 656  
DB 2100 CCTGTCTCTGTTTGAATGGAATTCGAACCGGGTACTTGCTCTCAAGAGTGGGTATTA 2159  
OY 656 nValAspArgIlyrIleAlaSerGluValCysGluAspSerIleThrValGlyMetValAr 676  
DB 2160 AGTGAAGAAAGTACATGCTCCGGAAGTGTGTGACAGATCATCCTGTGGGAAGCTGTA 2219  
OY 676 gHisGlnGlyLysIleMetTyrValGlyAspValArgSerValThrGlnLysHisIleG 696



Db 2220 GCATGAAGCCAGATCAAAATATGTCAATGACGTCGCGAAATATCAACGAAGAAATATGA 2299  
Qy 696 nGluTrpGlyProPheAspLeuValIleGlyGlySerProCysAsnAspLeuSerIleVa 716  
Db 2280 AGAGTGGGCGCCGTCGAGCTTGGTGATTTGGTGAAGCCCATGAATGATCTCTAACGT 2339  
Qy 716 lAsnProAlaArgLeuGlyLeuTyrcGluGlyThrGlyArgLeuPheGluPheTyAr 736  
Db 2340 CAATCTGCCCAAGGTTTATGAGGCACAGGAGGCTCTTCTCGAGTTTACCA 2399  
Qy 736 gLeuLeuHisAspAlaAArgProLysGluGlyAspAspArgProPheTrpLeuPheGl 756  
Db 2400 CTTGCTGAATTATACCCGCCCAAGGAGGCGGACCAACCGTCCATCTTCTGGATGTTTCA 2459  
Qy 756 uAsnValValAlaMetGlyValSerAspLysArgAspIleSerArgPheLeuGluSerAs 776  
Db 2460 CAATGTTGTGGCCATGAAGTGAATGACAGAAAGACATCTCAAGATTCTCGGATGTAA 2519  
Qy 776 nProValMetIleAspAlaLysGluValSerAlaAlaHisArgAlaArgTyPheTrpGl 796  
Db 2520 CCCAGTGTATGATCGATCGCATCAAGGTGTCTGCTGCTCACAGGCGCCGCTACTTCTGGG 2579  
Qy 796 yAsnLeuProGlyMetAsnArgProLeuAlaSerThrValAsnAspLysLeuGluLeuGl 816  
Db 2580 TAACTACCCGGAATGAACAGGCCCGTGGTCTCAAGAAATGATTAAGCTCGAGCTGCA 2639  
Qy 816 nGluCysLeuGluHisGlyArgIleAlaLysPheSerLysValArgThrIleThrAr 836  
Db 2640 GGAGTCTGGAGTTCAGTAGGACACAGCAAGTTAAAGAAAGTGCAGCAATAAACCAACCA 2699  
Qy 836 gSerAsnSerIleLysGlnGlyLysAspGlnHisPheProValPheMetAngLuLysGl 856  
Db 2700 GTCGAATCCATCAGACAGCGGCAAAACCAAGCTTTTCCCTGTAGTCATGAATGCAAGGA 2759  
Qy 856 uAspIleLeuTyrcGlyThrGluMetGluArgValPheGlyPheProValHisTyThrAs 876  
Db 2760 CGACGTTTGTGGTGCACGTAGCTGCAAGAGATCTTCGGCTTCCCTGCTCACTACACGGA 2819  
Qy 876 pValSerAsnMetSerArgLeuAlaArgGlnArgLeuLeuGlyArgSerTrpSerValPr 896  
Db 2820 CGTGTCCAACATGGGCGCGCGCGCTCAGAAAGCTGCTGGGAGGTCCTGGAGGTACC 2879  
Qy 896 oValIleArgHisLeuPheAlaProLeuLysGluTyrcPheAlaCys 911  
Db 2880 GGTCAATCAGACACCTGTTTGGCCCTTGAAGGACTACTTTGCCCTGT 2925

RESULT 11  
AF151973  
LOCUS  
DEFINITION Mus musculus DNA cytosine-specific methyltransferase isoform 5  
ACCESSION AF151973  
VERSION AF151973.1 GI:8347127  
KEYWORDS  
SOURCE Mus musculus.  
ORGANISM Mus musculus.  
REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 4223)  
Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and Shen,Y.  
TITLE Cloning of full-length Dnmt3b cDNA and its alternative splicing isoforms in mouse embryonic tissue  
JOURNAL  
REFERENCE Unpublished  
AUTHORS 2 (bases 1 to 4223)  
Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and Shen,Y.  
TITLE Direct Submission  
JOURNAL Submitted (17-MAY-1999) Department of Biochemistry and Molecular Biology, Institution of Basic Medical Sciences, Chinese Academy of Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005, P.R.China

FEATURES  
source  
Location/Qualifiers  
1..4223  
/organism="Mus musculus"  
/strain="KM"  
/db\_xref="taxon:10090"  
/dev\_stage="8-9 day old embryo"  
1..4223  
/gene="Dnmt3b"  
297..2876  
/gene="Dnmt3b"  
/function="methylates cytosine in DNA"  
/note="alternatively spliced"  
/codon\_start=1  
/product="DNA cytosine-specific methyltransferase isoform 5".  
/protein\_id="AAF74519.1"  
/db\_xref="GI:8347128"  
/translation="MKGDSRHLNEEGASCVEECIIVNGHFSDOSSDTKDAPSPV  
ALCTEPVCTPETGRSSRLSKREVSSLLNYTQDMTGDGDRDDEVDDGSDILMPK  
LTREKNDTRIRSESPAVRTHSGTSLSERQASPRITRQGRHHVQVEYVEFFPATR  
SRRRRASSASPSPSPASVDPMEEVTPKSVSTPSVDLSQDQDQDMTTQVQDSIY  
GDSYQDDKEFGIGDLVMGKIKGFWMPAMVSWKATSKQAMPMKRWQWFGDGKF  
SEISADKLVALGFSQHLNATFNKLYSRKAMYHTLEKARVAKTSSSEGESLED  
QKPMLEWAHGGFKPTGIGLKPKNQPVVNVKSRVSDSNLSPRRRNSRRRTTN  
DGAASESPPKRLKTNVSGYKDRGEDESERMASVEVTNNKGLDCLSCQKNPVS  
PPLVFGGLCQCRDRFELFYMYLDBDGYOYCTVCCGCEGRELLELCSNCCRCFCVEC  
LEVLVAGTAEADAKLOEPWSCYMLPQCHGVLRPRKDMNMLQDFFTDPLESEP  
PKLYPAIPAAXRRPIRVLSLFDGIATGYLVKELGIVKEKYLAEVCAESVAGTVKH  
EQKIFVNDVRKLTGKIEWGFPLDVIGSGPCNDLSNVNPARKGLYEGTGRLFFEFY  
HLUNTRPKGDNRPFFMFNVVAMVNDKKOISRFLACNPVIDAIVKSAHARY  
FVGNLPMNRPVMAKNDLEQLDCLFSEATKLKKVQITTKGNSIRQKNQLPFPV  
MNGKDDVLMCTELRIFGPPAHYTDVSNMGRGARQKLLGRSWSVPVIRHLFAPLKDYF  
ACE"  
BASE COUNT 1051 a 1084 c 1147 g 941 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 2,07e-80 Length: 4223  
Score: 2223.50 Matches: 450  
Percent Similarity: 59.81% Conservative: 123  
Best Local Similarity: 46.97% Mismatches: 240  
Query Match: 45.06% Indels: 145  
DB: 10 Gaps: 14  
US-09-720-086-7 (1-912) x AF151973 (1-4223)  
Qy 41 ProSerThrThraAlaArgLysValGlyArgProGlyArgLysArgLysHisProProVal 60  
Db 171 CCAAGTAAACGTAGCGAGCGATCGCGCGGAGATTGCGGAACCCGACACTCCGCGCG 230  
Qy 61 GluSerGlyAspThrProLys-----AspProAlaValIleSer 73  
Db 231 CCGCGCGCGAGGACCGCGCGGATCGCGCGCGGCTACAGCCAGCC-----TCA 284  
Qy 74 LysSerProSerMetAlaGlnAspSer-----GlyAlaSerGlu 86  
Db 285 CCACAGGAACAATGAAGGAGAGAGACATCTGAATGAAGAGAGGTCGACGCGG 344  
Qy 87 -----LeuLeuProAsnGlyAspLeuGluLysArgSerGluProGlnProGlu 102  
Db 345 TATGAGGAGTGCATTATCGTTAATGGAACTTCAGTACCAGCTCTCAGACACGAAGGAT 404  
Qy 103 GluGlySerProAlaGlyGlyGlnLysGlyGlyAlaProAlaGluGlyGluGlyAlaAla 122  
Db 405 GCTCCCTCACCC-----416  
Qy 123 GluThrLeuProGluAlaSerArgAlaValGluAsnGlyCysCysThrProLys----- 140  
Db 417 -----CCAGTCTGGAGGCAATCTGCACAGAGCGAGTCTGCACACACGAGACCAGA 467  
Qy 140 -----140  
Db 468 GGCGCGAGGTCAAGTCCCGGCTGTCTTAAGAGGAGGAGTCTCCAGCGCTTCTGAATTACAG 527



```

Db 2577 GAGCTGAGGACTGCTGAGGTTTCAGTAGGACGCAAGTAAAGAAAGTGCACACATA 2636
Qy 834 ThrThrArgSerAsnSerIleLysGlnGlyAspGlnHisPheProValPheMetAsn 853
Db 2637 ACCACCAAGTCAACTCCATCAGACAGGCAAAACAGAGCTTTCCCTGTAGTCATGAAT 2696
Qy 854 GluLysGluAspIleLeuTyrCysThrGluMetGluArgValPheGlyPheProValHis 873
Db 2697 GCCAAGGACGAGCTTTTGTGGTCACTGAGCTCGAAGGATCTTCGGCTTCCCTGCTCAC 2756
Qy 874 TyrThrAspValSerAsnMetSerArgLeuAlaArgGlnArgLeuGlyArgSerTyr 893
Db 2757 TACAGGAGCGTGTCAACATGGCCCGCGCCGCTCAGAAAGCTGCTGGGAGAGCTCTGG 2816
Qy 894 SerValProValIleArgHisLeuPheAlaProLeuLysGluTyrPheAlaCys 911
Db 2817 AGTGTACCGGTTCATCAGACACCTGTTTGCCCTTGAAGGACTACTTTTGCCTGT 2870

RESULT 12
AF068626 4195 bp mRNA linear ROD 06-DEC-1999
LOCUS Mus musculus DNA cytosine-5 methyltransferase 3B1 (Dnmt3b) mRNA,
DEFINITION alternatively spliced, complete cds.
ACCESSION AF068626
VERSION AF068626.2 GI:6449469
KEYWORDS
SOURCE Mus musculus.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4195)
AUTHORS Okano,M., Xie,S. and Li,E.
TITLE Cloning and characterization of a family of novel mammalian DNA
(cytosine-5) methyltransferases
JOURNAL Nat. Genet. 19 (3), 219-220 (1998)
MEDLINE 98324766
PUBMED 9662389
REFERENCE 2 (bases 1 to 4195)
AUTHORS Xie,S., Okano,M. and Li,E.
TITLE Direct Submission
JOURNAL Submitted (28-MAY-1998) CVRC, Mass. Gen. Hospital, 149 13th Street,
Charlestown, MA 02129, USA
REFERENCE 3 (bases 1 to 4195)
AUTHORS Okano,M., Chijiwa,T., Sasaki,H. and Li,E.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-1999) CVRC, Mass. Gen. Hospital, 149 13th Street,
Charlestown, MA 02129, USA
REMARK Sequence update by submitter
COMMENT On Nov 18, 1999 this sequence version replaced gi:3327979.
FEATURES
     source
         1..4195
             /organism="Mus musculus"
             /db_xref="taxon:10090"
             /chromosome="2"
             /map="84.0 CM"
             /gap="4195"
             /gene="Dnmt3b"
             /gene="Dnmt3b"
             /note="similar to EST sequences deposited in GenBank
             accession numbers AA116694, AA119979, AA17277, AA210568,
             AA407106, and AA575617"
             /map="2848"
             /gene="Dnmt3b"
             /function="de novo DNA methylation"
             /note="alternatively spliced product; contains Cys-rich
             region; C-terminal region is similar to corresponding
             region of DNA cytosine-5 methyltransferase 3A"
             /codon_start=1
             /product="DNA cytosine-5 methyltransferase 3B1"
             /protein_id="AAC40178.2"
             /db_xref="GI:6449470"

gene
misc_feature
CDS

```

```

/translation="MKGDSRHLNEERGASGYERCLIVNGNFSDOSSDTKDAPSPVLE
ALTEPVCPTETGRSSRLSREVSLLNYTQDMTGDGDRDDEVDGNGSDILMPK
LTPREKDTRESAPVTRHNSGTSLERQASPRITRQGRHHVQEVPEVFPAIR
SRRRASASSTPSPASVPMEEVTPKSVSTPSVLDSDQDGDMDTTQVDAESRD
GDSITYQDDKEFGIDLVNGKIFGFWPMVAVSWKATSKRQAMPQMWQFQDGRF
SSISADKLVALGFLSFQHNFLATLVSKYKAMVHTLEKARVAKGTSSSSGESLED
QLKPMLEWAHGGFKPTGI EGLKPKKQPVVNVKVRSSDSNLRPRRENKSRRTTN
DSAAESPPLKLTNSYGGKDRGDEESERMASEVTNNKGNLEDRCLSCQKKNPV
PHLEPGGLQCSCDRDFLELPYVYDEGYSYCTVCCGCEGRELILLCSTSCCEFCVE
LEVLVAGTAEADAKLOEPWSCYMLPQRCHVLRKRRKDMNMLQDPTTDDPLESEF
PKLYPAIPAAKRRPIRVLSLFDGATGYLVKELGIVKVEIKVISEVCAEYAVGVK
EQGIKYVNDVRKRTKNIIEWGPFDLVIGGSPNDLSNVNPARGLYSGTGRLEFFY
HLLNTPRKEGDNRPFFWPFNVAMVNDKDISRFLACNPVMDIALKXSAHARY
FWGNLPGMNRPMVMSKDLQDCLFESRTAKLKVQITTKNSIRQGNQLFPVY
MNGKDDVLWCTELERIFGFPAHYTDSVMGRGARQKLGRSWSVPIRHLFAPLKDYP
ACE"
BASE COUNT 1061 a 1059 c 1125 g 950 t
ORIGIN
Alignment Scores:
  2,59e-80      Length: 4195
Score: 2221.00   Matches: 456
Percent Similarity: 58.32%   Conservative: 126
Best Local Similarity: 45.69%   Mismatches: 256
Query Match: 45.01%   Indels: 160
DB: 10          Gaps: 15
US-09-720-086-7 (1-912) x AF068626 (1-4195)
Qy 16 AlaAlaGluArgGluGluAspArgLysAspGlyGluGlnGluGluProArgGlyLys 35
Db 23 GCGGCCCAAGTAACATGAGCGAGCATCGCGCGCGGATTCGGAACCC-----GAC 76
Qy 36 GluGluArgGlnGluProSerThrThrAlaArgLysValGlyArgProGlyArgLysArg 55
Db 77 ACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 136
Qy 56 LysHisPro---ProValGluSerGlyAspThrProLysAspProAlaValile----- 72
Db 137 CCTCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 196
Qy 73 -----SerLysSerProSerMet 78
Db 197 CCAGACGCGGGATCTCCCTCCCTCCATAGTCCTTGGGACCAATCCAGGCGCTTC 256
Qy 79 AlaGlnAsp-----SerGly 83
Db 257 TTTCAGAAACAATGAAGGAGACAGACACATCTGAATGAAGAGAGGGTCCAGCGG 316
Qy 84 AlaSerGlu---LeuLeuProAsnGlyAspLeuGlyLysArgSerGluProGlnProGlu 102
Db 317 TATGAGGAGTGCATTATCGTTAATGGGAACCTTCAGTCACCACTCCCTCAGACACGAGGAT 376
Qy 103 GluGlySerProAlaGlyGlyGlnLysGlyAlaProAlaGluGlyGluGlyAlaAla 122
Db 377 GCTCCCTCACCC-----388
Qy 123 GluThrLeuProGluAlaSerArgAlaValGluAsnGlyCysCysThrProLys----- 140
Db 389 -----CCAGCTTTGAGGCAATCTGCACAGAGCGAGTCTGCACACACAGACACCA 439
Qy 140 -----140
Db 440 GGCGCGCAGGTCAAGCTCCGGCTGTCTAAGAGGGAGGTCTCCAGCGCTTCTGAATTACAC 499
Qy 141 -----GluGlyArgGlyAlaPro 146
Db 500 CAGGACATCAGCAGGAGATGGAGACAGAGATGATGAAGTAGATGATGGGAATGGCTCTGT 559
Qy 147 AlaGluAlaGlyLysGluGlnLysGluThrAsnIleGluSerMetLysMetGluGlySer 166
Db 560 ATTCTAATGCCAAAGACTCACCCGTGAGACCAAGACACACAGGACGCGCTCTGAAAGCCG 619

```



```

Db 2669 GCGAAGGACGCGTTTGTGGTGCACTGAGCTCGAAGGATCTTCGGCTCCCTGCTCAC 2728
Qy 874 TyrThrAspValSerAsnMetSerArgLeuAlaArgGlnArgLeuLeuGlyArgSerTrp 893
Db 2729 TACACGAGCTGTCAACATGGCCCGCGCGCTCAGAAAGTGTCTGGGAGGCTCTGG 2788
Qy 894 SerValProValIleArgHisLeuPheAlaProLeuLysGluTyrPheAlaCys 911
Db 2789 AGTGTACCGGTATCAGACACCTGTGTGGCCCTTGAAGACTACTTTGCTGT 2842

RESULT 13
AF151969
LOCUS
DEFINITION
  AF151969 4338 bp mRNA linear ROD 08-JUN-2000
  Mus musculus DNA cytosine-specific methyltransferase isoform 1
  (Dnmt3b) mRNA, complete cds.
ACCESSION
  AF151969
VERSION
  AF151969.1 GI:8347117
SOURCE
  Mus musculus.
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  1 (bases 1 to 4338)
AUTHORS
  Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
  Shen,Y.
TITLE
  Cloning of full-length Dnmt3b cDNA and its alternative splicing
  isoforms in mouse embryonic tissue
JOURNAL
  Unpublished
REFERENCE
  2 (bases 1 to 4338)
  Yin,B., Chen,Y.T., Zhu,M., Luo,Y.J., Zhu,N., Xu,S.C., Wu,G.Y. and
  Shen,Y.
DIRECT SUBMISSION
  Submitted (17-MAY-1999) Department of Biochemistry and Molecular
  Biology, Institution of Basic Medical Sciences, Chinese Academy of
  Medical Sciences, Dong Dan San Tiao 5, Beijing City 100005,
  P.R.China
FEATURES
    source
      1..4338
        /organism="Mus musculus"
        /strain="KM"
        /db_xref="taxon:10090"
        /dev_stage="8-9 day old embryo"
      1..4338
        /gene="Dnmt3b"
        412..2991
        /gene="Dnmt3b"
        /function="methylates cytosine in DNA"
        /note="alternatively spliced"
        /codon_start=1
        /product="DNA cytosine-specific methyltransferase isoform
        1"
        /protein_id="AAF74515.1"
        /db_xref="GI:8347118"
        /translation="MKGDSRLHNEEGAGSYECIIVNGNFSDDSSDTKDAPSPVLE
        AICTPEVCTPTRGRSSRLSKREVSLNLTQDMTGDDRDDEDDGNGSDILMPK
        LTRKDTKTRKSPFAVRIRNSGTSSLERQASPRITGRGRHHRVQETPVEFPATR
        SRRRASSTPWSASPASVDPMEEVTPKSVSTPSVDLSODGDEGMDTTOVDAESIY
        GDSTEYQDDKEFGIDLVNGIKGFSWPMVYVSWKATSKRQMPGMWQVDFGDKF
        SEISADKLVALGFSQHFNLATFNKLVSKYKMYHTLEKARVAGTFFSSGESLED
        QLKPLMEWAHGPKPTGTIEGLKPKKQPVNKSVDNRLEPRRNRSRRTTN
        DSAASESPPKRLKNTSYGKDRGDEESRMASEVTNNKGNLEDRCLSCGKNPVS
        FHLPEGLQSCDRDLELFVYDEDDGVSYCTVCCGRELILCSNTSCRCFCVEE
        LEVLVAGTAEADAKLQEPSCVMCLPQCHGVLRERKDNMRLODFETDPLEEP
        PKLYPAIPAKRPIRLVSLPDGIATVLVLKELGIVKEKYIASEVCASIIAGTVKH
        EGQIKYVNDVRKTIYKNIIEEWGPFDLVIGGSPCNLSNVNPAKRGVETGRUFFEY
        HLLNTYKPKEDNPPFFWMFNVMVANKNDKDISRLACPNPMIDAIKVSAAHARY
        FWGLPGNRPVMAKNDKLEDFRSTAKLKKVQTTIKSNIRIQGNQLFPV
        MNGKDDVLWCTELERIFGPPAHYTDVSNMGRGARQKLLGRSWSVPVIRHLFAPLKYF
        ACE"
BASE COUNT 1073 a 1124 c 1173 g 968 t
ORIGIN
Alignment Scores:

```

```

Pred: No.: 2,93e-80 Length: 4338
Score: 2220.00 Matches: 462
Percent Similarity: 57.87% Conservative: 126
Best Local Similarity: 45.47% Mismatches: 253
Query Match: 44.98% Indels: 175
DB: 10 Gaps: 17

US-09-720-086-7 (1-912) x AF151969 (1-4338)

Qy 2 ProAlaMetProSer---SerGlyProGlyAspThrSerSerSer-----AlaAlaGlu 18
Db 147 CGCGAGCGCCGGGGTTAAGCGGCCCAAGTAACGTAGCGAGCGATCGGCGCCGAGAT 206
Qy 19 -ArgGluGluAspArgLysAspGlyGluGluGluGluProArgGlyLysGluGluAr 38
Db 207 TCGCGAACCAGC-----ACTCCGCG 227
Qy 38 gGlnGluProSerThrAlaArgLysValGlyArgProGlyArgLysArgLysHisPr 58
Db 228 CGCGCCGCGCGCGAGCCCGCGCGCGATCGCGCGCCGCTACAGCCAGCCTCACGA 287
Qy 58 o---ProValGluSerGlyAspThrProLysAspProAlaValIle-----72
Db 288 CAGGCCCGCTGAGGCTTGTGCCAGACCTTGGAAACCTCAGGTATATACCTTTCCAGACGC 347
Qy 73 -----SerLysSerProSerMetAlaGlnAs 81
Db 348 GGGATCTCCCTCCCCCATCCATAGTGCCTTGGGACCAATCCAGGGCCTTCTTTTCAGGA 407
Qy 81 p-----SerGlyAlaSerGl 86
Db 408 AACAAATGAAGGGAGACAGACACATCTGAATGAAGAAGAGGGTGCACGGGTATGAGGA 467
Qy 86 u---LeuLeuProAsnGlyAspLeuGluLysArgSerGluProGlnProGluGluGlySe 105
Db 468 GTGCATATTCGTTAATGGGAACCTTCAGTGACCACTCTTCAGACAGGAAGGATGCTCCCTC 527
Qy 105 rProAlaGlyGlyGlnLysGlyGlyAlaProAlaGluGlyGluGlyAlaAlaGluThrLe 125
Db 528 ACC-----531
Qy 125 uProGluAlaSerArgAlaValGluAenGlyCysCysThrProLys-----140
Db 532 -CCAGTCTTGGAGCAATCTGCACAGAGCCAGTCTGCACACAGAGACAGAGCGCGCAG 590
Qy 140 -----140
Db 591 GTCAAGTCCCGCTGTCTTAAGAGGGAGGTCTCCAGCTTCTGAATTCACAGCAGGACAT 650
Qy 141 -----GluGlyArgGlyAlaProAlaGluAl 149
Db 651 GACAGGAGATGGAGACAGAGATGATGAAGTAGATGATGGGAATGCTCTGATATTCTAAT 710
Qy 149 agLysGluGlnLysGluThrAsnIleGluSerMetLysMetGluGlySerArgGlyAr 169
Db 711 GCCAAGCTTACCGGTGAGACCAAGGACACAGGACCGCTCTGAAAGCCCGGCTGTCCG 770
Qy 169 gLeuArgGlyGlyLeuGlyTrpGluSerSerLeuArgGlnArgProMetProArgLeuTh 189
Db 771 AACCCGACATAGCAATGGGACCTCCAGCTTGGAGAGGCAAGAGCCCTCCCCAGAAATCAC 830
Qy 189 r-----PheGlnAlaGlyAspProTyrTril 198
Db 831 CGAGGTCCGCGAGGCGCCACCATGTGCGAGGAGTACCTCTGTGGAGTTTCCGGTACCAG 890
Qy 198 eSerLysArgLysArgAspGluTrpLeuAlaArgTrpLysArgGluAlaGluLysAl 218
Db 891 GTCTCGAGACGTCGA-----906
Qy 218 aLysValIleAlaGlyMetAsnAlaValGluGluAenGlnGlyProGlyGluSerGlnly 238
Db 907 -----GCATCGTCTTCAGCAAGACGCGCCATGCTATCCCTGCCAGCGTGCATT 956

```

QY 238 svaIgluIuaIaSerProProAlaValIgluInIProThraSproAlaSerProThra 258  
 Db 957 CATGAAAGAGTGCACCTTAAGAGCGTCAATCCCA-----TCAGTTGACTT 1004  
 QY 258 IAlaThrProgluProValGlySerAspAlaGlyAspIlyAsnaAlaThrIlysaIaG1 278  
 Db 1005 GAGCAGAGATGAGATCAGAGAGGTATGATACCAACAGGTGATGCAGAGCATATA 1064  
 QY 278 yAspAspIuProgluTyrgIuAspGlyArgGlyPheGlyIleGlyIleuValTyrG1 298  
 Db 1065 TGGAGACACGACAGATATACAGATGATTAAGAGTTTGAAATAGGTGACTCTGTGGG 1124  
 QY 298 yLysLeuArgGlyPheSerTrpTrpProGlyArgIleValSerTrpTrpMetThrGlyAr 318  
 Db 1125 AAGATCAAGGCTTCTCTGTGGCTCCATGTGTGTCTGGAAGCCACTTCCAA 1184  
 QY 318 gSerArgAlaAlaGluGlyThrArgTrpValMetTrpPheGlyAspGlyLysPheSerVa 338  
 Db 1185 GCGACAGCCATGCCGGAATGCCGTGGGTACAGTGTGTGTGTGTGTGTGTGTGTGA 1244  
 QY 338 IValCyseValGlyLysLeuMetProLysSerSerPheCyseSerAlaPheHileGlnAlaTh 358  
 Db 1245 GATCTCTCTGCAAACTGCTGCTCTGGGCTGTTCAGCCAGCACTTTAATCTGCTAC 1304  
 QY 358 rTyAsnLysGlnProMetTyArgLysAlaIleTyrgIuValIleuGlnValAlaSerSe 378  
 Db 1305 CTTCATTAAGCTGGTTCTTATAGAGGCCATGTACACACTCTGGAGAAAGCCAGGTT 1364  
 QY 378 rArgAlaGlyLysLeuPheProValCyseHisAspSerAspIuSerAspThrAlaLysAl 398  
 Db 1365 TCGAGCTGGCAAGACCTTC-----TCCAGCAGCTCTGAGAGGTC 1403  
 QY 398 aValGluValGlnAsnLysProMetIleGluTrpAlaLeuGlyGlyPheGlnProSerG1 418  
 Db 1404 ACTGGAGACCAAGCTGAAGCCATGCTGAGAGTGGCCAGGCTGCTTCAAGCTTACTGG 1463  
 QY 418 yProLysGlyLeuGluPro----- 424  
 Db 1464 GATGAGAGGCCCTCAAAACCAACAAAGAGCAACAGTGTAAATAGTCAGAGGTCGCTCG 1523  
 QY 425 -----ProGluGluGlyLysAsnProTyrgLysGluValTyrgTh 437  
 Db 1524 TTCAGACAGTAGAAGACTTAAGAACCCAGAGAGCGGAGAACAAAGCTCGAAGAGCGCAAC 1583  
 QY 437 rAspMetTrpValGluProGluAlaAlaIleTyrgAlaProProProAlaLysLysPr 457  
 Db 1584 CAAT-----GACTGCTGCTCTTCTGAGTCCCCCCCAACCAAGGCGCTCAA 1628  
 QY 457 oArgLysSerThrAlaGlyLysProLysValIlysgIuIleIleAspGluArgThrArg1 477  
 Db 1629 GACAAATAGCTATGGCGGAGAGACCGAGGGAG-----GATGAGAGAGCCGAGA 1679  
 QY 477 uArgLeuValTyrgIuValArgGlnLysCyseArgAsnIleGluAspIleCyseIleSerCy 497  
 Db 1680 ACGGATGCTTCTGAAGTCAACCAACAAAGGCGCAATCTGGAAGACCGCTGTCTCTG 1739  
 QY 497 sGlySerLeuAsnValThrIleuGlnHisProLeuPheValIleGlyIleMetCyseGlnAsnCy 517  
 Db 1740 TGGAAAGAAAGAACCTGTGCTCTTCCACCCCTCTTGAGGGGTGGCTGTCTGAGAGTGG 1799  
 QY 517 sLysAsnCyPheLeuGluCyseAlaTyrgIuTyrgAspAspGlyTyrgIuInsertyTyCy 537  
 Db 1800 CCGGAGATGCTTCTCTGAGCTCTTCTACATGTATGATGAGAGACCGCTATACAGCTCTACTG 1859  
 QY 537 sThrIleCyseGlyGlyArgGluValLeuMetCyseGlyAsnaAsnAsnCyseArgCy 557  
 Db 1860 CACCGTGTGCTGTGAAGGCGGTGTAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1919  
 QY 557 sPheCyValGluCyseValAspLeuLeuValGlyProGlyAlaAlaGlnAlaIleTy 577  
 Db 1920 CTTCTGT 1979  
 QY 577 sGluAspProTrpAsnCyTyrgMetCyseGlyHisLysGlyThrTyrgIleuLeuArgAr 597

Db 1980 GCAGAAACCTGGAGCTGCTATATGTGCTCCCTCCAGCGCTGCCATGGGCTCTCCACG 2039  
 QY 597 gArgGluAspTrpProSerArgLeuGlnMetPhePheAlaAsnaHisAsp---GlnG1 616  
 Db 2040 CAGGAAAGATTGGAAATGCGGCTCGCAAGACTTCTTCACTACTGATCTGACTCGGAAGA 2099  
 QY 616 uPheAspProProLysValTyrgProProValProAlaGluLysArgLysProIleArgVa 636  
 Db 2100 ATTGAGCCACCCCAAGTTGTATACCAAGCAATCTCTGCAGCCAAAGAGGCCATTTAGAGT 2159  
 QY 636 lLeuSerLeuPheAspGlyIleAlaThrGlyLeuLeuValLeuLysAspLeuGlyIleG1 656  
 Db 2160 CTTCTCTGTTTGTATGGAATTGCAACGGGGTACTTGCTGTCTCAAGAGACTTGGGTATTTAA 2219  
 QY 656 nValAspArgTyrgIleAlaSerGluValCyseGluAspSerIleThrValIleMetValAr 676  
 Db 2220 AGTGGAAAGATCATTTGCCCTCCGAAGTCTGTGCAGAGTCCATCCGTGTGGAACTGTTAA 2279  
 QY 676 gHisGlnGlyLysIleMetTyrgValGlyAspValArgSerValThrGlnLysHisIleG1 696  
 Db 2280 GCATGAAGGCCAGATCAAAATATGTCATGACGCTCCGGAATAATCACAGAAATAATTTGA 2339  
 QY 696 nGluTrpGlyProPheAspLeuValIleGlyLysProCyseAsnaAspLeuSerIleVa 716  
 Db 2340 AGAGTGGGGCCCGTTCGACTTGATGATTTGTGTGAAGCCCATGATGATCTCTTAACGT 2399  
 QY 716 lAsnProAlaArgLysGlyLeuTyrgIuGlyLysArgLysPhePheGluPheTyrgAr 736  
 Db 2400 CAATCTCCGCCCAAGGTTTATATGAGGGCACAGGAAGGCTCTTCTGAGTTTATCCA 2459  
 QY 736 gLeuLeuHisAspAlaArgProLysGluGlyAspAspArgProPhePheTyrgPheG1 756  
 Db 2460 CTTCGTGAATTTATACCGCCCAAGAGGGCGACAAACCGTCATCTTCTGTGATGTTTCCA 2519  
 QY 756 uAsnValValAlaMetGlyValSerAspLysArgAspIleSerArgPheLeuGluSerAs 776  
 Db 2520 GAATGTTCTGGCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2579  
 QY 776 nProValMetIleAspAlaLysGluValSerAlaAlaHisArgAlaArgTyrgPheTrpG1 796  
 Db 2580 CCCAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2639  
 QY 796 yAsnLeuProGlyLysAsnArgProLeuAlaSerThrValAsnaAspLysLeuGluG1 816  
 Db 2640 TAACTTACCCGGAATAAAGAGGCGCGGTGATGCTTCAAGATATGATGATGATGATGATGAT 2699  
 QY 816 nGluCyseLeuGlnHisGlyArgIleAlaLysPheSerLysValArgThrIleThrTrAr 836  
 Db 2700 GGAATGCTGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2759  
 QY 836 gSerAsnSerIleLysGlnGlyLysAspGlnHisPheProValPheMetAsnGluLysG1 856  
 Db 2760 GTCGAATCCCATCAGACAGGCGCAAAACAGCTTTCCTGTAGTCATGAATGGCAAGA 2819  
 QY 856 uAspIleLeuTyrgTyrgThrGluMetGluArgValPheGlyPheProValHisTyrgTrAs 876  
 Db 2820 CGAGCTTGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2879  
 QY 876 pValSerAsnMetSerArgLeuAlaArgGlnArgLeuLeuGlyArgSerTrpSerValPr 896  
 Db 2880 CGTGTCCCAACATGGGCGCGCGCGCGCTGCAAGAGTGTGTGTGTGTGTGTGTGTGTGTGT 2939  
 QY 896 oValIleArgHisLeuPheAlaProLeuLysGluTyrgPheAlaCyS 911  
 Db 2940 GGTATCAGACACCTGTGTGCTTCCCTTGAAGAGACTTACTTGTCTGT 2985  
 RESULT 14  
 AF331857 4335 bp mRNA linear PRI 02-JAN-2002  
 LOCUS Homo sapiens DNA cytosine methyltransferase 3 beta (DNMT3B) mRNA,  
 DEFINITION complete cds.  
 ACCESSION AF331857



```

VERSION AF331857.1 GI:18033254
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 4335)
  AUTHORS Ni, J., Pradhan, S. and Roberts, R.J.
  TITLE Cloning, expression and characterization of human DNMT3 genes
  JOURNAL unpublished
REFERENCE
  2 (bases 1 to 4335)
  AUTHORS Ni, J., Pradhan, S. and Roberts, R.J.
  TITLE Direct Submission
  JOURNAL Submitted (22-DEC-2000) New England Biolabs, 32 Tozer Road,
  Beverly, MA 01915, USA
FEATURES
    source
        1..4335
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
    gene
        1..4335
        /gene="DNMT3B"
    CDS
        322..2883
        /gene="DNMT3B"
        /codon_start=1
        /product="DNA cytosine methyltransferase 3 beta"
        /protein_id="AA157040.1"
        /translation="MKGDTIRHLNGEDAGREDISILVNGACSDQSSDPPILEAIRTP
        EIRGRSSRSKRREVPSSLLSYTQDLTGDDGEDGSDTPVPMKLFRETRRSSPA
        VTRNNSSRSKRREVPSSLLSYTQDLTGDDGEDGSDTPVPMKLFRETRRSSPA
        PPSYLLIDLTDDTDTHTGTPQSSPTPYARLAQDSQGGMESPPQVADSGDGSSEYQ
        DKDFGIGDLVNGKIKGFSWPMVSVKATSKRQAMSGMRVQWFDGDKFSEVSADK
        LVALGPSOHNELATFNKLVSVKMYHAKLEKARVRAGKTFPSSPGDSLEDDQKPMLE
        WAHGFPGTIGLEKPNPTOPVVKNSKVRAGSRKLESRYENKTRRTTRADDATSQY
        CPAPRLKNTNNGKDRGDEDSREOMASDVANNKSLSDGCLSCGRKNPVSPHPLP
        EGGLCQICRDRLEFLFYMDDDGYQSYCTVCEGRELLELLCSNFTSCRCFCVEGLEVLV
        GTGAARAKLQEPWSCVCLPQRCQVLRKRDQNVRLQAFSTDGTLEAEPKLYPA
        IPAARRPIRVLSPDGIATGYLKEIGIKGVKYVASEVEESIAVGTVKBHGNIRKY
        VDNRRITKNIIEEMGPFDLVIGGSPNDLSNVNPAKGLYEGLTGRFFEFHLLNYS
        RPKGPDPRFFMFENYVAMVKYGRDIDSRFLFCNPFMDIAIKVSAARHARYFWGNLP
        GMRPVIASKNDKLEQDCLYNIKAKKKVQITITTKNSIKQGNQLFPVWNGKED
        VLWCTELERIFGFPVHVTVDSNMGRGARKLLGRSWSVPVIRHLFAPLKDYFACE"
BASE COUNT 1045 a 1172 c 1167 g 951 t
ORIGIN
Alignment Scores:
Pred. No.: 6.69e-80 Length: 4335
Score: 2211.00 Matches: 477
Percent Similarity: 59.54% Conservative: 122
Best Local Similarity: 47.42% Mismatches: 263
Query Match: 44.80% Indels: 147
DB: 9 Gaps: 23

US-09-720-086-7 (1-912) x AF331857 (1-4335)
QY 2 ProAlaMetProSerSerGlyProGlyAspThrSerSerAlaAlaGluArgGluGlu 21
Db 8 CCGCGTCCCGCG---TCCGCGCCCGCGCGCTTCTCCGACGAGCTGTCTCC-CGCGTCCGC 63
QY 22 AspArgLysAspGlyGluGluGluProArgGlyLysGluGluArg---GlnGlu 40
Db 64 GGCCGC-----AGCCCGCGTGGAGCTCCGAGCGCCCGCCGACG 102
QY 41 ProSerThrAlaArgLysValGlyArgProGlyArgLysArgLysHisProProVal 60
Db 103 GACGGGACCGGCTCCCTCGCGGTGGCGGAGCGCGGCGAACGCTGCCCGCGCGGACGCG 162
QY 61 GluSerGlyAspThrProLysAspPro---AlaValIleSerLysSerProSerMetAla 79
Db 163 CTGGGGTTAAGTGGCCCAAGTAACCTAGCTCGGCGATCGGCGCGGAGATTCCGAGCC 222
QY 80 Gln-----AspSerGlyAlaSerGluLeuLeuProAsnGlyAspLeuGlu 94
Db 111 Gln-----AspSerGlyAlaSerGluLeuLeuProAsnGlyAspLeuGlu 94

```

```

Db 223 CAGCGCGCTTGCACGCGCGCGCCAGCGCGCTCCCGCCAGCCAGCCCGCGGCTCCGC 282
QY 95 LysArgSerGluProGlnPro-----
Db 283 CGCCACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 101
QY 102 -----GluGluGlySerProAlaGlyGlyGlnLys-----GlyGly 113
Db 343 CTCATATGGAGAGAG-----GACGCGCGCGGGGAGGAGAGACTCGATCTCTGCTAACCGG 396
QY 114 AlaProAlaGluGlyGluGlyAlaAlaGluThrLeuProGluAlaSerArgAlaValGlu 133
Db 397 GCGTGCAGCGACGAGTCTCCGACTCCGACCTCCGACCTCCGAGGCTATCCGC-----447
QY 134 AsnGlyCysCysThrProLys---GluGlyArgGlyAlaProAlaGluAlaGlyLysGlu 152
Db 448 -----ACCCCGGAGATCAGAGCGCGGAGATCAAGCTCCGACTCTCCAGAGG 495
QY 153 Gln-----
Db 496 GAGGTGTCCAGTCTGCTAAGCTACACAGGACTTGACAGCGGATGGCGAGCGGGAAGAT 555
QY 154 -----LysGluThrAsnIleGluSer 160
Db 556 GGGGATGGCTCTGACACCCCGCTATGCCAAAGCTCTTCCGGGAAACAGGACTCGTTCA 615
QY 161 MetLysMetGluGlySerArgGlyArgLeuArgGlyGlyLeuGlyTrpGluSerSerLeu 180
Db 616 -----GAAAGCCAGCTGTCCGAACTCGAATAACAAACAGTGTCTCCAGCGCGGAG 666
QY 181 ArgGlnArgProMetProArgLeuThrPheGlnAlaGlyAspProTyrTyrIleSerLys 200
Db 667 AGGCACAGCGCTTCCCGACGTTCCACCC--GAGGCGCGC-----703
QY 201 ArgLysArgAspGluTrpLeuAla---ArgTrp-----210
Db 704 AGGCGCGCAACCATGTGGACGATCCCGCTGGAGTTCCTCCGGGTACCGAGTCCCTGAGAC 763
QY 211 -----LysArgGluAlaGluLysLysAlaLysValIleAlaGlyMetAsnAlaVal 227
Db 764 GCGCGGCAACAGCATCGGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 823
QY 228 GluGluAsnGlnGlyProGlyGluSerGlnLysValGluGluAlaSerProProAlaVal 247
Db 824 TCGACCTCAG-----ACGACACAGAGGACACACATGGGACGCGCCCGAGACGACA 874
QY 248 GlnGlnProThrAspProAlaSerProThrValAlaThr-----ThrProGlu 263
Db 875 GTACCCCTTACG---CCCGCTAGCCCGACGACGACGAGCGGCGGCGCATGGAGTCCCGCG 931
QY 264 ProValGlySerAspAlaGlyAspLysAsnAlaThrLysAlaGlyAspAspGluProGlu 283
Db 932 AG-GTGGAGCGACAGTGGAGAT-----GGAGACAGTTCAGAG 969
QY 284 TyrGluAspGlyArgGlyPheGlyIleGlyGluLeuValTrpGlyLysLeuArgGlyPhe 303
Db 970 TATCAGGATGGGAGAGTTCGATATAGGGGACCTCGTGTGGGAAAGATCAAGGGCTTC 1029
QY 304 SerTrpTrpProGlyArgIleValSerTrpTrpMetThrGlyArgSerArgAlaAlaGlu 323
Db 1030 TCCTGTGGCGCCCATGGTGTCTTGAAGGCCACCTTCAAGCGACAGGCTATGTCT 1089
QY 324 GlyThrArgTrpValMetTrpPheGlyAspGlyLysPheSerValValCysValGluLys 343
Db 1090 GGATCGCGTGGGTTCAGTGTGTGGCGATGGCAAGTTCCTCCGAGGCTCTCTGCACACAAA 1149
QY 344 LeuMetProLysSerSerPheCysSerAlaPheHisGlnAlaThrTyrAsnLysGlnPro 363
Db 1150 CTGGTGGCACTGGGGGTGTTCAGCCAGCACTTAAATTTGGCCACCTTCAATAAGCTCGTC 1209
QY 364 MetTyrArgLysAlaIleTyrGluValLeuGlnValAlaSerSerArgAlaGlyLysLeu 383
Db 1210 TCCTATCGAAGACCATGATACCATGTCTCTGGAGNAAGTAGGTGCGAGCTGGCAGACC 1269

```



```

OY 384 PheProValCysHisAspSerAspGluSerAspThrAlaValAlaGluValAsn 403
Db 1270 TTCCCC-----AGCAGCCCTGGAGACTCATTTGGAGCAGCAGCTG 1308
OY 404 LysProMetIleGluTrpAlaLeuGlyPheGlnProSerGlyProLysGlyLeuGlu 423
Db 1309 AAGCCCATGTTGAGTGGCCCGAGGGCTTCAGGCCCTGAGATCGAGGGCTCAAA 1368
OY 424 ProProGluGlu----- 427
Db 1369 CCCAACAACACGCAACCGAGTGTATTAAGTCGAGAGGTGCTGTCAGAGCAGTAGGAAA 1428
OY 428 ---GluLysAsnProTyrLysGluValTyrThrAspMetTrpValGluProGluAla 446
Db 1429 TTAGATCAAGGAATAACGAAACAGACTCGAAGACGACAGCTGACGACCTGACCAAC 1488
OY 47 Ala---TyrAlaProProProProAlaValLysProArgLysSerThrAlaGluLysPro 465
Db 1489 TCTACTACTGCCCCCGACCCAGCCCTCAAGCAAAATTGCTATTAACAACGGCAAAAGAC 1548
OY 466 LysValLysGluIleLeuAspGluArgThrArgGluArgLeuValTyrGluValArgGln 485
Db 1549 CGAGGGAGTGA-----GATCAGAGCCGAGAACAAATGGCTTCAGATGTTGCCAAC 1599
OY 486 LysCysArgAsnIleGluAspIleCysIleSerCysGlySerLeuAsnValThrLeuGlu 505
Db 1600 AACAGAGCAGCTCGAAGAAATGCGTGTGTCTGTGGCAGGAAAAACCCCGTCTCTTC 1659
OY 506 HisProLeuPheValGlyGlyMetCysGlnAsnCysLysAsnCysPheLeuGluCysAla 525
Db 1660 CACCTCTCTTTGAGGGGGGCTGTGTGACACATGCCGGGATCGCTTCTTGAGCTGTT 1719
OY 526 TyrGlnTyrAspAspAspGlyTyrGlnSerTyrCysThrIleCysCysGlyGlyArgGlu 545
Db 1720 TACATGATGAATGACATGCTATCAGCTTACTGCACTGTGTGCTGCGAGGGCCGAGAG 1779
OY 546 ValLeuMetCysGlyLysAsnAsnAsnCysCysArgCysPheCysValGluCysValAspLeu 565
Db 1780 CTGCTGCTTTGACAGCAACAGAGCTGCTGCGGTGTGTGTGTGTGTGTGTGTGTGTGT 1839
OY 566 LeuValGlyProGlyAlaAlaGlnAlaAlaIleLysGluAspProTyrAsnCysTyrMet 585
Db 1840 CTGTGGGACAGGACAGCAGCGGCCAGAGCTTCAGAGCCCTCGAGCTGTACATG 1899
OY 586 CysGlyHisLysGlyThrTyrGlyLeuLeuArgArgArgGluAspTyrProSerArgLeu 605
Db 1900 TGTCTCCGCGCAGCGCTCATGCGCTCTGCGGCGCGGAGGAGCTGGAACGTGCGCTG 1959
OY 606 GlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProLysValTyrProPro 625
Db 1960 CAGGCTTCTTCAACAGTGAACAGCGGGCTTGAATATGAAAGCCCAAGCTGTACCTGCC 2019
OY 626 ValProAlaGluLysArgLysProLysArgValLeuSerLeuPheAspGlyIleAlaThr 645
Db 2020 ATTCCCGCAGCCGAGAGCGGCCCATTCAGATCTCTGATTTGATGATGATGATGATGATG 2079
OY 646 GlyLeuLeuValLeuLysAspLeuGlyIleGlnValAspArgTyrIleAlaSerGlyVal 665
Db 2080 GCGTACCTAGTCTCAAGAGATGGGCAATTAAGTAGGAAAGTACGTGCTTCTGAAGTG 2139
OY 666 CysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIleMetTyrValGly 685
Db 2140 TGTGAGAGTCCATGCTGTGTTGGAACCGTAGAGCAGAGGGAATATCAATACGTGAAC 2199
OY 686 AspValArgSerValThrGlnLysHisIleGlnGluTrpGlyProPheAspLeuValIle 705
Db 2200 GACGTGAGGAACATCAACAAGAAATATTTGAAGAATGGGCCCATTTGATCTGTGATTT 2259
OY 706 GlyLysSerProCysAsnAspLeuSerIleValAsnProAlaArgLysGlyLeuTyrGlu 725
Db 2260 GCGGGAAGCCCATGCAAGATCTCTCAAAATGTGAATCCAGCCAGAAAGCCGTGATAG 2319

```

```

OY 726 GlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAlaArgProLysGlu 745
Db 2320 GGTACAGCGCGGCTCTTCTTGAATTTTACACCTGCTGATTTACTACAGCCCAAGAG 2379
OY 746 GlyAspAspArgProPhePheTrpLeuPheGluAsnValAlaIleMetGlyLysSerAsp 765
Db 2380 GGTATGACCGGCGCTTCTTCTGATGTTTGAATGTTTGAAGCATGAGCATGAGTTGGCGAC 2439
OY 766 LysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLysGluVal 785
Db 2440 AAGAGGACATCTCAGCGTTCCTGAGTGAATTCAGTGAATGATGATGATGATGATGAT 2499
OY 786 SerIleAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMetAsnArgProLeu 805
Db 2500 TCTCTGCTCAGAGGCGCCGATCTCTGGGGCACTTACCGGGATGAACAGCCCGTG 2559
OY 806 AlaSerThrValAsnAspLysLeuGluLeuGlnGluCysLeuGluHisGlyArgIleAla 825
Db 2560 ATACATCAAGAAATATTAATCAAGCTGACAGGACTGCTTGGAATTAAGATAGGCC 2619
OY 826 LysPheSerLysValArgTrpIleThrThrArgSerAsnSerIleLysGlnGlyLysAsp 845
Db 2620 AAGTTAAAGAAATACAGACAAATTAACCAAGTCAAGTCAATCAATCAATCAATCAATCA 2679
OY 846 GlnHisPheProValPheMetAsnGluLysGluAspIleLeuTrpCysThrGluMetGlu 865
Db 2680 CAACTTTCCCTGTTGTGATGAATGGCAAGAGATGTTTGTGTGCTGACTGAGCTGAA 2739
OY 866 ArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeuAlaArg 885
Db 2740 AGGATCTTGGCTTCTTCTGTGACATACACAGAGCTGTCAACATGGGCGGTGTGCCGC 2799
OY 886 GlnArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAlaProLeu 905
Db 2800 CAGAACTGCTGGAGAGTCTTGGAGCTGCTGTATCCGACACCTTTCGCCCTCTG 2859
OY 906 LysGluTyrPheAlaCys 911
Db 2860 AAGCACTACTTTCATGT 2877

RESULT 15
LOCUS AF176228 4267 bp mRNA linear PRI 26-OCT-1999
DEFINITION Homo sapiens DNA cytosine-5 methyltransferase 3B (DNMT3B) mRNA,
complete cds.
ACCESSION AF176228
VERSION AF176228.1 GI:6118091
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4267)
XU,G.-L., Bestor,T.H., Bourc'his,D., Hsieh,C.-L., Tommerup,N.,
Bugge,M., Hulten,M., Qu,X., Russo,J.J. and Viegas-Pequignot,E.
Chromosomes in a DNA methyltransferase gene
Nature (1999) in press
2 (bases 1 to 4267)
XU,G.-L. and Bestor,T.H.
Direct Submision
Submitted (06-AUG-1999) Genetics and Development, Columbia
University, 701 West 168 St., New York, NY 10032, USA
FEATURES
source
1..4267
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/map="20q11-q13"
1..4267
/gene="DNMT3B"
265..2802
/gene="DNMT3B"

```

/note="mutated in human chromosome instability and immunodeficiency disease known as ICF syndrome; contains alternative 5' exon 1B"

```

/codon start=1

```

```

/sequence_3B"
/product="DNA cytosine-5 methyltransferase 3B"

```

```
/product= BSA Cytochrome  
/protein_id="AAF04015.1
```

```
/db_xref="GI:6118092"
```

[illegible]

BASE COUNT	1069 a	1098 c	1147 q	953 f
	RI	FG	FP	VH
	I	D	V	S
	N	M	G	R
	G	A	R	Q
	K	L	L	G
	R	S		

**BASE CO  
ORIGIN**

Alignment Scores:

Alignment Scores:			
Pred. No.:	1.37e-79	Length:	4267
Score:	2203.00	Matches:	455
Percent Similarity:	60.75%	Conservative:	110
Best Local Similarity:	48.92%	Mismatches:	221
Query Match:	44.64%	Indels:	145
DB:	9	Gaps:	17

2000

US-09-720-086-7 (1-912) x AF176228 (1-4267)

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

Qy 6 SerSerGlyProGlyAspThrSerSer-----

[illegible]

US-09-720-086-7 (1-912) x AF176228 (1-4267)

6	SerSerGlyProGlyAspThrSerSer	-----SerAla 16
368	TCAACGGGCGCTGCAGCACCAGTCTCCGACTCGCCCCCAATCTCTGGAGGCTATCCGCA	427
17	AlaGluArgGluGluAspArgLysAspGlyGluGluGlnGluGluProArgGlyLys---	35
428	CCCCGGAGATCMAG-----GCCGAAGATCAAGCTCGGACTCTCCAAGAGGGAGGTG	481
36	-----GluGluArgGlnGluProSerThrThrAlaAlaArgLysValGlyArg 50	
482	CCAGTCTCTAAGCTACACACAGACTTGACAGCGGATGCGCAGCGGGGAAGATGGG---	538
51	ProGlyArgLysArgLysHisProProValGluSerGlyAspThrProLysAspProAla 70	
539	-----ATGGCTCTGCACACCCCACTCATGTCAAC 565	
71	ValIleSerLysSerProSerMetAlaGlnAspSerGlyAlaSerGluLeuLeuProAsn 90	
566	AGCTCTTCCGGGAACACAGGACTCTGTTTCAGAAAGCCAGCTGTCCGAACTC-----	616
91	GlyAspLeuGluLysArgSerGluProGlnProGluGluGlySerProAlaGlyGlyGln 110	
617	-----GAAATAACAACAGTGTCTCCAGCCGG-----	643
111	LysGlyGlyAlaProAlaGluGlyGluGlyAlaAlaGluThrLeuProGluAlaSerArg 130	
644	-----AGAGGCACAGGCCTTCCCAACGTTCCACCCGAGCGCCGCAGG 685	
131	AlaValGluAsnGlyCysCysThrProLysGluGlyArg---GlyAlaProAlaGluAla 149	
686	CCCCAACCATGTGCAGCAGTCCCCCGTGGAGTTCGGGCTACAGGTCTCTGAGACCGC 745	
150	GlyLysGluGlnLysGluThrAsnIleGluSerMetLysMetGluGlySerArgGlyArg 169	
746	GGCAACACAGCATCGCAGAA-----	766
170	LeuArgGlyGlyLeuGlyTrpGluSerSerLeuArgGlnArgProMetProArgLeuThr 189	

QY 542 G1yglYArgGluValIleuMetCysglYAsnAsnAsnCysCysArgCysPheCysValGlu 561  
Db 1687 GAGGCGCCGAGAGCTGCTGCTTGGACGACACGAGCTGCTGCCGGTGTCTTCTGTGTGAG 1746  
QY 562 CysValAspLeuLeuValGlyProGlyAlaAlaGlnAlaAlaIleYsgLAspProTyr 581  
Db 1747 TGCCTGAGGTGCTGTGGGACACAGCGCCGAGGCCCAAGCTTCAGGAGCCCTGG 1806  
QY 582 AsnCysTyrMetCysGlyHisLysGlyTyrTyrGlyLeuLeuArgArgGluAspTyr 601  
Db 1807 AGCTGTACATGTGTCTCCCGAGCGCTGTATGCGCTGCGCGGCGGAGAGACTGG 1866  
QY 602 ProSerArgLeuGlnMetPhePheAlaAsnAsnHisAspGlnGluPheAspProProlys 621  
Db 1867 AACGTGCGCTGAGCGCTTCTTCCACAGTGACACGCGGCTTGAATATGAAAGCCCCAAG 1926  
QY 622 ValTyrProProValProAlaGluLysArgLysProIleArgValIleuSerLeuPheAsp 641  
Db 1927 CTGTACCTCCGCAATCCCGAGCGCCGAGGCCCAATTCAGCTCTGTCAATGTGTGAT 1986  
QY 642 GlyIleAlaThrGlyLeuLeuValIleuLysAspLeuGlyIleGlnValAspArgTyrIle 661  
Db 1987 GGCATCGGACAGGCTACCTAGTCTCANAAGAGTTGGCATAAAGTAGGAAAGTACGTC 2046  
QY 662 AlaSerGluValCysGluAspSerIleThrValGlyMetValArgHisGlnGlyLysIle 681  
Db 2047 GCTTCTGAAGTGTGTGAGAGCTCATGTCTGTGGAACCGTGAAGCACGAGGGGAATATC 2106  
QY 682 MetTyrValGlyAspValArgSerValThrGlnLysHisIleGlnGluTyrGlyProPhe 701  
Db 2107 AAATACGTGAACGACGTGAGGAACATCACAAAGAAATATTGAAGATGGGCCCATTT 2166  
QY 702 AspLeuValIleGlyGlySerProCysAsnAspLeuSerIleValAsnProAlaArgLys 721  
Db 2167 GACTTGGTATGCGCGGAGGCCCATGACATCGATCTCTAAATGTGAATCCAGCCAGGAA 2226  
QY 722 GlyLeuTyrGluGlyThrGlyArgLeuPhePheGluPheTyrArgLeuLeuHisAspAla 741  
Db 2227 GGCCTGTATGAGGTATACAGCGCGCTCTTCTGAAATTTACACCTGTGAATTACTCA 2286  
QY 742 ArgProLysGluGlyAspAspArgProPhePheTyrLeuPheGluAsnValAlaIleMet 761  
Db 2287 CGCCCCAAGAGGTGTATGACCGCGCTCTTCTGAAATTTACACCTGTGAATTACTCA 2346  
QY 762 GlyValSerAspLysArgAspIleSerArgPheLeuGluSerAsnProValMetIleAsp 781  
Db 2347 AAGTTGGCGACAGAGGAGACATCTCACGCTTCTGAGTGAATTCAGTGAATGATTGAT 2406  
QY 782 AlaLysGluValSerAlaAlaHisArgAlaArgTyrPheTyrPheTyrPheTyrPheTyr 801  
Db 2407 GCCATCAAGTTTCTGCTCTCACAGGCCCGATCTTCTGGGGCAACCTAACCCGGGATG 2466  
QY 802 AsnArgProLeuAlaSerThrValAsnAspLysLeuGlnGluCysLeuGluHis 821  
Db 2467 AACAGGCCCTGTATGATCAAAAGATGATTAACCTCGAGCTGCAGGACTGCTTGAATAC 2526  
QY 822 GlyArgIleAlaLysPheSerLysValArgThrIleThrThrArgSerAsnSerIleLys 841  
Db 2527 AATAGATGCGCAAGTAAAGAAAGTACAGACAAATTAACCAACCAAGTCGAACTCGATCAA 2586  
QY 842 GlnGlyLysAspGlnHisPheProValPheMetAsnGlnLysGluAspIleLeuTyrCys 861  
Db 2587 CAGGGGAAAAACCACTTTTCTCTGTGTGATGATGCAAAAGAGATGTTTGTGTGTC 2646  
QY 862 ThrGluMetGluArgValPheGlyPheProValHisTyrThrAspValSerAsnMetSer 881  
Db 2647 ACTGAGCTCGAAGATCTTGGCTTTCTGTGTGATGATGATGATGATGATGATGATGATG 2706  
QY 882 ArgLeuAlaArgLysLeuGluArgSerTyrSerValProValIleArgHisLeu 901  
Db 2707 CGTGTGCGCCGACAGAGCTGTGGAGAGTCTCGAGCGTGCCTGTCAATCCGACACCTC 2766

QY 902 PheAlaProLeuLysGluTyrPheAlaCys 911  
Db 2767 TTCGCCCTCTGAAAGACTACTTTCATGT 2796  
Search completed: November 22, 2002, 10:57:33  
Job time : 5471.49 secs

**THIS PAGE BLANK (USPTO)**

28	40.2	1.0	635	5	PCT-US94-05354-35	Sequence 35, Appl
29	40.2	1.0	1100	3	US-09-248-335-53	Sequence 53, Appl
30	40.2	1.0	1466	4	US-08-984-919A-10	Sequence 10, Appl
c 31	40.2	1.0	1466	4	US-08-984-919A-12	Sequence 12, Appl
32	40.2	1.0	1472	4	US-08-781-420-10	Sequence 10, Appl
c 33	40.2	1.0	1472	4	US-08-781-420-12	Sequence 12, Appl
34	40.2	1.0	1472	4	US-08-874-102-10	Sequence 10, Appl
c 35	40.2	1.0	1472	4	US-08-874-102-12	Sequence 12, Appl
36	40.2	1.0	1472	4	US-09-006-595A-10	Sequence 10, Appl
c 37	40.2	1.0	1472	4	US-09-006-595A-12	Sequence 12, Appl
38	40.2	1.0	1493	6	5340934-5	Patent No. 5340934
39	40.2	1.0	1875	4	US-08-984-919A-46	Sequence 46, Appl
c 40	40.2	1.0	1875	4	US-08-984-919A-48	Sequence 48, Appl
41	40.2	1.0	1881	4	US-08-874-102-46	Sequence 46, Appl
c 42	40.2	1.0	1881	4	US-08-874-102-48	Sequence 48, Appl
43	40	1.0	2852	3	US-09-027-137-2	Sequence 2, Appli
44	40	1.0	2852	4	US-09-344-441-2	Sequence 2, Appli
45	39.8	0.9	337	2	US-09-032-684-11	Sequence 11, Appl

## ALIGNMENTS

## RESULT 1

US-09-276-531-47

; Sequence 47, Application US/09276531

; Patent No. 6183968

## ; GENERAL INFORMATION:

; APPLICANT: Bandman, Olga

; APPLICANT: Lal, Preeti

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Yue, Henry

; APPLICANT: Reddy, Roopa

; APPLICANT: Guegler, Karl J.

; APPLICANT: Baughn, Mariah R.

; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING

; TITLE OF INVENTION: RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION

; NUMBER OF SEQUENCES: 134

## ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

; STREET: 3174 PORTER DRIVE

; CITY: PALO ALTO

; STATE: CALIFORNIA

; COUNTRY: USA

; ZIP: 94304

## ; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2

## ; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/276,531

; FILING DATE: Herewith

## ; CLASSIFICATION:

## ; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/079,677

; FILING DATE: March 27, 1998

## ; CLASSIFICATION:

## ; ATTORNEY/AGENT INFORMATION:

; NAME: Lynn E. Murry, Ph.D.

; REGISTRATION NUMBER: 42,918

; REFERENCE/DOCKET NUMBER: PA-0008 US

## ; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (650) 855-0555

; TELEFAX: (650) 845-4166

## ; INFORMATION FOR SEQ ID NO: 47:

## ; SEQUENCE CHARACTERISTICS:

; LENGTH: 2077 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

## ; IMMEDIATE SOURCE:

; LIBRARY: TESTTUT02

THIS PAGE BLANK (USPTO)

CLONE: 1271435  
US-09-276-531-47

Query Match	13.1%	Score 548.4;	DB 4;	Length 2077;
Best Local Similarity	71.4%	Prod. No. 1,86-144;		
Matches 888; Conservative	0;	Mismatches 306;	Indels 49;	Gaps 11;

2295 GGCCTCTTCTTGAGTTTAACTTCTGTGAATTATACCGGCCCCAAGAGGGCCACAC 2354  
 65 GGCTCTTCTTGAAATTTT-CCACCTGCTGAAATTACTCACGCCCCCAAGAGGGGTGATGACC 123  
 2355 GTTCATTCTTTTGATGTTCGAGATGTGTGCCCATGAAAGTGAATGACAACAAAGACA 2414  
 124 GGNCGTCTCTTGATGTGTTGAGAAATGTTGATGNCCTGTCGCCAATTCCGACAGGGGACA 183  
 2415 TCTCAAGATTCCTGGCACTGTAACCCAGATGATCGATGCGATCAAGGGTGTCTGCTC 2474  
 Db TCTCACGGTCTCTGGAGTGTAAATCCAGTGATATGATCCATTAAAGTTTCTGCTGCTC 243  
 2 ACAGGGCCCCGTACTTCTGGGGTAAACCAACCGGAATGAACAGGCCCCGTGATGCTTCAA 2534  
 Db ACAGGGCCCCGTAATCTCTGGGGCAACCTAACCCGGATGAAACAGGCCCGTGATGATCAA 303  
 2535 AGAATGATTAAGCTCGAGCTGACGACCTGCTGGAGTTCACTAGAGACAGCAAGTTAAAGA 2594  
 Db AGAATGATTAACCTCGACCTGACGACCTGCTGGAGTTCACTAGAGATGCCAATGAAGA 363  
 2595 AAGTGCACACATTAACCAACCAAGTGCATCTCATACAGACAGGGCAAAAACCACTTTCC 2654  
 Db AAGTACAGCAATTAACCAACCAAGTGCATCTCATACAGAGGGAAAAAACCACTTTTCC 423  
 2655 CTGTAGTCATGAATGCGAAGACGACGTTTGTGTGTCTGAGCTTGAAGAATCTTCC 2714  
 Db CTGTGTCAATGAATGCGAAGAAAGATGTTTGTGTGTGTGACACTGACCTGAAAGATCTTTG 483  
 2715 GCTCCCTGCTCACTACACGACGCTGTCCAAATGCGGCCGCGCCGCTGCAGAACTGC 2774  
 Db GCTTCTCTGCTGACTTACACAGACTGTCCAAATGCGGCCGCTGTGCGCCGACAGAACTGC 543  
 2775 TGGCAGAGTCTCGAGTGTACCGGTATCAGACACTGTTGCCCCCTTGAAAGACTACT 2834  
 Db TGGAAAGTCTCGAGAGCTGCTGTATCCGACACTCTTGGCCCCCTGAAGGACTACT 603  
 2835 TTGCGTGTGAATAGTTCTTACCCAGAGACTGGGAGCTCTG-----GTACAGCCAGTG 2887  
 604 TTGCATGTGAATAGTTCCAGCCAGGGGCCAAGCCCACTGGGATGTGTGGCAGAGCCAGGA 663  
 664 CCCAGGAGGTGTGATTTCTCGAAGGCA--TCCCCAAGGCCGCTCTTCTCAGCTGTGTG 721  
 2948 GGCCCTCATCATCTGTAACCTTCTCTGCTAGTGGAGACAGAGCTCTTGCCCC 3007  
 722 GTATAT---ACGTGTACTCAATTCCTCTTGCTCAGTGGGGCAGAGCCACTGATCTC 777  
 3008 TTGCAGGGGAGCCCC--GGTCTTCCCTCGTGTGACAGCTCAGACCTGGGCTGTAGAG 3065  
 778 TTGCAGGGGTAGCCTGAGGTGCGCGCTCTTGTGCACAAATACAGACTGTGCTTGAGAG 837  
 3066 TAGCCCCGCGATGTGCTCATGTTCTTTAACCTGAAACTTTAAACTTGAATAGGTAGT 3125  
 838 CAGCCTTAACACGCTGCTCATTTTCTTCTCTTAATAACTTTAAACTTGAAATAGGTAGC 897  
 3126 AAGATGGC--TTCTTTTAACCTTCCTGAGTTTATCACTC--AGAAGTATGCTTAAGAT 3188  
 898 AAGCTGGTTTTTTTTTTTCCCTTCCCTGAGTCAACACTCAGAGAAACATGTCTAAGAT 957  
 3181 ACCAAAAAACAACAAACAAACAGAAACAAACAAACAAACAAACAAACCTCAACAGCTCTCTTA 3248  
 958 ACCAAACACACAGTGGCGACAGCTCTCCAACTACAGGTTAATGTGAAAAATCATCAAA 10117  
 3241 GTACTCAGGTATGCTGCAAAATCACTGAAATTTGTTTAAATGAACCGGTGCTCA 3300

[illegible]



THIS PAGE BLANK (USPTO)

13	158.5	3.2	14255	5	PCT-US94-04496-1	Sequence 1, Appl
14	157	3.2	4201	1	US-08-080-255-4	Sequence 4, Appl
15	157	3.2	4201	3	US-08-465-713-4	Sequence 4, Appl
16	157	3.2	4201	5	PCT-US93-05857-4	Sequence 4, Appl
17	153	3.1	11220	4	US-09-105-537-32	Sequence 32, Appl
18	153	3.1	36778	4	US-09-105-537-5	Sequence 5, Appl
19	153	3.1	38506	3	US-09-320-878-19	Sequence 19, Appl
20	152.5	3.1	8458	1	US-07-945-283-1	Sequence 1, Appl
21	151.5	3.1	4852	1	US-07-853-913-3	Sequence 3, Appl
22	151.5	3.1	6755	4	US-08-931-999-4	Sequence 4, Appl
23	150.5	3.1	5267	3	US-08-976-255-2	Sequence 2, Appl
24	150	3.1	2355	4	US-08-913-159-12	Sequence 12, Appl
25	149	3.0	13842	4	US-09-130-242-1	Sequence 1, Appl
26	149	3.0	36778	4	US-09-105-537-30	Sequence 30, Appl
27	149	3.0	38506	3	US-09-105-537-5	Sequence 5, Appl
28	149	3.0	38506	3	US-09-320-878-19	Sequence 19, Appl
29	148.5	3.0	41280	2	US-08-804-227C-1	Sequence 1, Appl
30	148	3.0	1556	2	US-08-405-175A-10	Sequence 10, Appl
31	148	3.0	9551	1	US-08-056-200-93	Sequence 93, Appl
32	147.5	3.0	9551	2	US-08-800-644-93	Sequence 93, Appl
33	147.5	3.0	2115	3	US-09-032-365A-12	Sequence 12, Appl
34	147.5	3.0	2116	1	US-08-701-380-1	Sequence 1, Appl
35	147	3.0	2680	4	US-09-063-035-1	Sequence 1, Appl
36	147	3.0	15872	4	US-09-105-537-1	Sequence 1, Appl
37	146	3.0	1926	4	US-09-249-585A-2	Sequence 2, Appl
38	146	3.0	2580	4	US-09-050-863-2	Sequence 2, Appl
39	146	3.0	5452	2	US-09-359-081-2	Sequence 2, Appl
40	146	3.0	9600	4	US-09-130-114-1	Sequence 1, Appl
41	146	3.0	9600	4	US-08-910-647-1	Sequence 1, Appl
42	146	3.0	10596	1	US-09-620-925-1	Sequence 1, Appl
43	146	3.0	10596	1	US-07-884-811-15	Sequence 15, Appl
44	146	3.0	10596	1	US-07-885-971-15	Sequence 15, Appl
45	146	3.0	10596	1	US-08-087-783A-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1  
US-09-276-531-47  
Sequence 47, Application US/09276531  
Patent No. 6183968  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Lal, Preeti  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Yue, Henry  
APPLICANT: Reddy, Roopa  
APPLICANT: Guegler, Karl J.  
APPLICANT: Baughn, Mariah R.  
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF GENES ENCODING  
NUMBER OF SEQUENCES: 134  
TITLE OF INVENTION: RECEPTORS AND PROTEINS ASSOCIATED WITH CELL PROLIFERATION  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/276,531  
FILING DATE: Herewith  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/079,677  
FILING DATE: March 27, 1998  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

NAME: Lynn E. Murry, Ph.D.	REGISTRATION NUMBER: 42,918
REFERENCE/DOCKET NUMBER: PA-0008 US	TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555	TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 47:	SEQUENCE CHARACTERISTICS:
LENGTH: 2077 base pairs	TYPE: nucleic acid
STRANDEDNESS: single	TOPOLOGY: linear
IMMEDIATE SOURCE:	LIBRARY: TESTTUT02
CLONE: 1271435	US-09-276-531-47
Alignment Scores:	
Pred. No.: 3 99e-52	Length: 2077
Score: 748.00	Matches: 151
Percent Similarity: 81.95%	Conservative: 17
Best Local Similarity: 73.66%	Mismatches: 36
Query Match: 15.24%	Indels: 4
DB: 4	Gaps: 1
US-09-720-086-5 (1-908) X US-09-276-531-47 (1-2077)	
QY 703 GlySerProGlyAsnAspLeuSerIleValAsnProAlaArgGlyLeuTyGlyGly 722	
DB 3 GCGGCGNANTGCACACATCTNMAAT-GTGAATCA-GCCAGGAAGCGTGTAGAGGAC 60	
QY 723 ThrGlyArgLeuPheGluPheGlyThrArgLeuLeuHisAspAlaArgProGlyGly 742	
DB 61 AGG--CGGCTCTTCTCGAATTTC-CACGCTGCAATTAACCGCCCAAGAGGCT 116	
QY 743 AspAspArgProPhePhePheGluAsnValValAlaMetGlyValSerAspLys 762	
DB 117 GATGACCGGNCCTTCTTGAGATCTTGAGATGTTGTAGNCTGCGGATTCGCGACA 176	
QY 763 ArgAspIleSerArgPheLeuGluSerAsnProValMetIleAspAlaLysGlyValSer 782	
DB 177 GGGGACATCTCACGGTCTCTGGAGGTGTAATCCAGTATGATGATGCCATCAAGTTTCT 236	
QY 783 AlaAlaHisArgAlaArgTyrPheTrpGlyAsnLeuProGlyMetAsnArgProLeuAla 802	
DB 237 GCTGCTACAGGCGCCGATATCTTGGGGCAACTTACCCGGATGAACAGCCCGGATGA 256	
QY 803 SerThrValAsnAspLysLeuGluGluGluGluGluGluGluGluGluGluGluGluGlu 822	
DB 297 GCATCAAGATGATTAACCTCGNCTCAGACCTCTTGAAATCAATAGATAGATAGATAG 356	
QY 823 PheSerLysValArgThrIleThrThrArgSerAsnSerIleLysGluGluGluGluGlu 842	
DB 357 TTAAGAAAGTACAGACATTAACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 416	
QY 843 HisPheProValPheMetAsnGluGluGluGluGluGluGluGluGluGluGluGluGlu 862	
DB 417 CTTTCCCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 476	
QY 863 ValPheGlyPheProValHisTyrThrAspValSerAsnMetSerArgLeuAlaArgGln 882	
DB 477 NTCTTGGCTTCTCTGTCATACACAGACGTCTTCAACATGAGGCGCTGCTGCTGCTG 536	
QY 883 ArgLeuLeuGlyArgSerTrpSerValProValIleArgHisLeuPheAlaProLeuLys 902	
DB 537 AAGCTGCTGGGAAGTCTCTGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 596	
QY 903 GluTyrPheAlaCys 907	
DB 597 GACTACTTGACATCT 611	
RESULT 2	
US-08-924-345-1	

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 22, 2002, 08:17:42 ; Search time 543.737 Seconds

(without alignments)  
17780.325 Million cell updates/sec

Title: US-09-720-086-3

Perfect score: 4293

Sequence: 1 gccgcgcgcacccggcgcgc.....taaaaaaaaaaaaaaa 4293

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0  
2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4283.4	99.8	4416	21	AAZ37097	DNA encoding de no
2	2870.2	66.9	4192	21	AAZ37095	DNA encoding de no
3	2752.2	64.1	2938	24	ABL90391	Human polynucleoti
4	764.4	17.8	4145	21	AAZ37098	DNA encoding de no
5	760.4	17.7	4195	21	AAZ37096	DNA encoding de no
6	613	14.3	709	24	ABQ57540	Human colon cancer
7	548.2	12.8	1060	22	AAH33860	Human colon cancer
8	504.6	11.8	622	22	AAH00271	Human reproductive
9	492.2	11.5	3017	22	AAH14009	Human CDNA sequenc

10	344.2	8.0	711	23	AA586165	DNA encoding novel
11	324.6	7.6	2077	22	AA80537	Receptor #25 parti
12	285.2	6.6	301	16	AA721884	Human gene signatu
13	273	6.4	273	24	ABN97389	Gene #3887 used to
14	273	6.4	273	24	AB64238	Stomach cancer rel
15	254.6	5.9	546	22	ABA20255	Human nervous syst
16	234.8	5.5	283	22	ABA13731	Human nervous syst
17	182.4	4.2	2951	23	AA586166	DNA encoding novel
18	173	4.0	411	22	AA185785	Human polynucleoti
19	171.4	4.0	1705	22	AA82964	Human DNA methyltr
20	171.4	4.0	3002	21	AACT7455	Human ORFX ORF3010
21	126.2	2.9	448	22	ABA43755	Human breast cell
22	126.2	2.9	448	22	ABA54212	Human foetal liver
23	126.2	2.9	448	22	ABA23961	Probe #2427 for ge
24	126.2	2.9	448	22	AAK02490	Human brain expres
25	126.2	2.9	448	22	AAK27928	Human bone marrow
26	126.2	2.9	448	22	AAI12506	Probe #2439 for ge
27	126.2	2.9	448	22	AAI33860	Probe #2546 used t
28	126.2	2.9	448	22	AAI02414	Probe #2405 used t
29	126.2	2.9	448	22	AB502382	Human genome-deriv
30	107.4	2.5	762	22	AAH05250	Human CDNA clone (
31	86	2.0	237	24	AB177340	Human ovarian canc
32	86	2.0	573	24	AB182861	Human ovarian canc
33	78.2	1.8	430	22	AA185778	Human polynucleoti
34	75.2	1.8	456	22	AA180464	Human polynucleoti
35	69.6	1.6	326	23	AB161126	Human prostate exp
36	69.2	1.6	585	23	ABV58085	Human prostate exp
37	68.2	1.6	10732	21	AA10594	Gene encoding a su
38	66.8	1.6	7434	24	AA028386	Human chemically t
39	66	1.5	6486	24	AB067050	Human angiogenesis
40	65.8	1.5	40862	24	AB134072	Human immune syste
41	65.4	1.5	14615	22	AA546704	Tumour suppressor
42	65.4	1.5	14798	24	AB133032	Human immune syste
43	65.2	1.5	432	23	ABV58007	Human prostate exp
44	64.8	1.5	420	23	ABV58336	Human prostate exp
45	64.6	1.5	6359	24	ABK39945	Human chemically p

#### ALIGNMENTS

RESULT 1	
AAZ37097	
ID	AAZ37097 standard; DNA; 4416 BP.
AC	AAZ37097;
XX	
DT	27-MAR-2000 (first entry)
XX	
DE	DNA encoding de novo DNA cytosine methyltransferase DNMT3A.
XX	
KW	De novo DNA cytosine methyltransferase; DNMT3A; neoplastic disorder;
KW	carcinoma; sarcoma; leukaemia; DNA methylation; ss.
XX	
OS	Homo sapiens.
XX	
FX	
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	353..3091
FT	/*tag= a
FT	/product= "de novo DNA cytosine methyltransferase
XX	DNMT3A"
XX	
XX	
XX	WO9967397-A1.
XX	
XX	
XX	29-DEC-1999.
XX	
XX	25-JUN-1999; 99WO-US14373.
XX	
XX	25-JUN-1998; 98US-0090906.
XX	24-JUL-1998; 98US-0093993.
XX	
XX	(GHEO ) GEN HOSPITAL CORP.



1681 GCAGAGTCCGGGAACTTGAAGACATCTCCTGCGGAGCCCAATGTTACCTT 1740  
1804 GCAGAAAGTCCGGGAACTTGAAGACATCTCCTGCGGAGCCCAATGTTACCTT 1863  
1741 GGAACACCCCTCTTCTGTTGAGGAAATGCAAACTGCAAGAACTCTTCTGAGTG 1800  
1864 GGAACACCCCTCTTCTGTTGAGGAAATGCAAACTGCAAGAACTCTTCTGAGTG 1923  
1801 TGGTACACAGTACGACGACGAGCTACAGTCTTACCTGACCACTCTGCGGAGCCG 1860  
1924 TGGTACACAGTACGACGACGAGCTACAGTCTTACCTGACCACTCTGCGGAGCCG 1983  
1861 TGAAGTCTCATGTGCGGAAACAACACTGCTGAGGTGCTTTTGGCGGAGTGTGGA 1920  
1984 TGAAGTCTCATGTGCGGAAACAACACTGCTGAGGTGCTTTTGGCGGAGTGTGGA 2043  
1921 CCTCTTGGTGGGCGCGGGGCTGCGGAGCCATTAAAGAAAGCCCTGGAACCTGCTA 1980  
2044 CCTCTTGGTGGGCGCGGGGCTGCGGAGCCATTAAAGAAAGCCCTGGAACCTGCTA 2103  
1981 CATGTGCGGGAACAAGGATACCTACGAGCTGCTGCGGCGGAGAGAGCTGCGCCCTCCG 2040  
2104 CATGTGCGGGAACAAGGATACCTACGAGCTGCTGCGGCGGAGAGAGCTGCGCCCTCCG 2163  
2041 GCTCCAGATGTTCTTCTGCTAATTAACAAGACGAGAAATTTGACCTCCAAAGGTTTACC 2100  
2164 GCTCCAGATGTTCTTCTGCTAATTAACAAGACGAGAAATTTGACCTCCAAAGGTTTACC 2223  
2101 ACCGTGCCAGCTGAGAAAGAAAGCCCATCCGAGGTGCTGCTCTTCTTGTAGGAATGCG 2160  
2224 ACCGTGCCAGCTGAGAAAGAAAGCCCATCCGAGGTGCTGCTCTTCTTGTAGGAATGCG 2283  
2161 TACAGGCTCTGCTGCTGAAGGCTTGGGCACTTCAAGGTGGAACGCTACATTCCTCGGA 2220  
2284 TACAGGCTCTGCTGCTGAAGGCTTGGGCACTTCAAGGTGGAACGCTACATTCCTCGGA 2343  
2221 GGTGTGAGAGACTCATCAAGGTGGGCTGCGGACACAGGGAAGATCATGTACGT 2280  
2344 GGTGTGAGAGACTCATCAAGGTGGGCTGCGGACACAGGGAAGATCATGTACGT 2403  
2281 CGGAGACGTCCGAGCTGACACAGAAAGATATCCAGAGTGGGGCCATTGATCTGCT 2340  
2404 CGGAGACGTCCGAGCTGACACAGAAAGATATCCAGAGTGGGGCCATTGATCTGCT 2463  
2341 GATTGGGGGAGTCCCTGCAATGACCTCTCAATGCAACCTCTGCAAGGGCCCTCA 2400  
2464 GATTGGGGGAGTCCCTGCAATGACCTCTCAATGCAACCTCTGCAAGGGCCCTCA 2523  
2401 CGAGGGCACTGGCCGGCTCTTCTTGAATTTACCGGCTCTGATGATGCGCGGCCAA 2460  
2524 CGAGGGCACTGGCCGGCTCTTCTTGAATTTACCGGCTCTGATGATGCGCGGCCAA 2583  
2461 GAGAGGAGATGATCGCCCTTCTTCTGAGCTCTTGAATGATGAGGCGCTTGA 2520  
2584 GAGAGGAGATGATCGCCCTTCTTCTGAGCTCTTGAATGATGAGGCGCTTGA 2643  
2521 TGAACAAGAGGACATCTCGGATTTCTGAGTCCAACCTCTGATGATGATGCAAGA 2580  
2644 TGAACAAGAGGACATCTCGGATTTCTGAGTCCAACCTCTGATGATGATGCAAGA 2703  
2581 AGTGTAGCTGACACAGGGCCGCTACTTCTGGGGTAACTTCCGGATATGAACGGCC 2640  
2704 AGTGTAGCTGACACAGGGCCGCTACTTCTGGGGTAACTTCCGGATATGAACGGCC 2763  
2641 GTTGGCATCCACTGTGAATGATAGCTGAGCTGACAGAGTGTGAGCATGACAGAT 2700  
2764 GTTGGCATCCACTGTGAATGATAGCTGAGCTGACAGAGTGTGAGCATGACAGAT 2823  
2701 AGCCAAAGTTCAGCAAAGTGAAGCACTTACTAGAGGTCAAACCTCCATTAAGCAGGGCAA 2760  
2824 AGCCAAAGTTCAGCAAAGTGAAGCACTTACTAGAGGTCAAACCTCCATTAAGCAGGGCAA 2883

2761 AGACCAACATTTCTCTCTCTGTAAGTGAAGAAAGAGACATCTTATGTGCACTGAAT 2820  
2884 AGACCAACATTTCTCTCTCTGTAAGTGAAGAAAGAGACATCTTATGTGCACTGAAT 2943  
2821 GGAAGGGTATTTGGTTTCCAGTCCACTATATGACGCTCCAAACATGAGCGGCTTGGC 2880  
2944 GGAAGGGTATTTGGTTTCCAGTCCACTATATGACGCTCCAAACATGAGCGGCTTGGC 3003  
2881 GAGGCAAGACTGCTGGGCGGCTATGAGCGTGCATGATCCGACCTTCTGCTCC 2940  
3004 GAGGCAAGACTGCTGGGCGGCTATGAGCGTGCATGATCCGACCTTCTGCTCC 3063  
2941 GCTGAAGAGATTTTGGCTGTGTGTAAGGACATGGGGGCAAACTGAGTGAAGCA 3000  
3064 GCTGAAGAGATTTTGGCTGTGTGTAAGGACATGGGGGCAAACTGAGTGAAGCA 3123  
3001 AAGTTAAACAAACAAACAAACAAACAAACATTAATAACACCAAGAACATGAGATGG 3060  
3124 AAGTTAAACAAACAAACAAACAAACAAACATTAATAACACCAAGAACATGAGATGG 3183  
3061 AGAAGATATCAGCACCCGAGAGAAAGAAAGAAATTTAAACAAACCAAGAGCGGA 3120  
3184 AGAAGATATCAGCACCCGAGAGAAAGAAAGAAATTTAAACAAACCAAGAGCGGA 3243  
3121 AATACCGAGGCTTGTGCTTGGGAAAGGTTGGAACATCATCTCTGATTTTCAATG 3180  
3244 AATACCGAGGCTTGTGCTTGGGAAAGGTTGGAACATCATCTCTGATTTTCAATG 3303  
3181 TATTTCTCAGTCTATTAATAAAACAAACCAAGCTCCCTCCCTCCCTCCCTTCCCTT 3240  
3304 TATTTCTCAGTCTATTAATAAAACAAACCAAGCTCCCTCCCTCCCTCCCTTCCCTT 3363  
3241 TTTTGGTGGTGAAGCTTTTATTTTCTACTCTTCTTCAAGGGGTTTTCTGTTGTTGGG 3300  
3364 TTTTGGTGGTGAAGCTTTTATTTTCTACTCTTCTTCAAGGGGTTTTCTGTTGTTGGG 3423  
3301 TTTTGGTGGTGGTGAAGCTTTTATTTTCTACTCTTCTTCAAGGGGTTTTCTGTTGTTGGG 3360  
3424 TTTTGGTGGTGGTGAAGCTTTTATTTTCTACTCTTCTTCAAGGGGTTTTCTGTTGTTGGG 3483  
3361 ATAGTAACAATACCTTGCAGAGAAAGGTGGAGAGAGAGAAAGAAATTTTAA 3420  
3484 ATAGTAACAATACCTTGCAGAGAAAGGTGGAGAGAGAGAGAAAGAAATTTTAA 3543  
3421 GAAATCTATATTTGGGTT 3480  
3544 GAAATCTATATTTGGGTT 3603  
3481 TTTTACTATATATCTTTTCTTGTGTCTAGCTGATCAGATAGAGACAAAGCAG 3540  
3604 TTTTACTATATATCTTTTCTTGTGTCTAGCTGATCAGATAGAGACAAAGCAG 3663  
3541 GGGACGAAAGAGAGACACTGAGCGGCGAGCATTTCCCTCCAGCACTGCTGCT 3600  
3664 GGGACGAAAGAGAGACACTGAGCGGCGAGCATTTCCCTCCAGCACTGCTGCT 3723  
3601 GCCAGCACTTCTGTGTGACGCAAAACAGAACCCAGTTAGCAGAGGAGAGAGAA 3660  
3724 GCCAGCACTTCTGTGTGACGCAAAACAGAACCCAGTTAGCAGAGGAGAGAGAA 3783  
3661 CCACACAAGACATTTTCTTACAGTATTTCAAGTGTCTACCAACAGAAACCTTGAAGA 3720  
3784 CCACACAAGACATTTTCTTACAGTATTTCAAGTGTCTACCAACAGAAACCTTGAAGA 3843  
3721 AATCAGTTCTAGAAAGCGCTGTACTCTTGTTTACAGTTTATATATATGATAGATA 3780  
3844 AATCAGTTCTAGAAAGCGCTGTACTCTTGTTTACAGTTTATATATATGATAGATA 3903  
3781 TGAGAT 3840  
3904 TGAGAT 3963  
3841 TTCAAAACAGCAGATGAGTAAACAAACATCAGCTTCCAGCTTCTGCGCAAGGGTTT 3900

Db 3964 TTCAACAGCGAGATGAGTAAACATCATGCTTCCAGCTTGCCCTTCTGCGAAAGGGTTT 4023  
Qy 3901 CACCAAGGATGGAGAAAGGAGACAGCTTGCAGATGGCGGTTCTCAGGTGGGCTCTTC 3960  
Db 4024 CACCAAGGATGGAGAAAGGAGACAGCTTGCAGATGGCGGTTCTCAGGTGGGCTCTTC 4083  
Qy 3961 CCCTTGGTTGTAAACGAGTGAAGGAGGAGAACTTGGAGCCAGGTTCTCCCTGCCAAAA 4020  
Db 4084 CCCTTGGTTGTAAACGAGTGAAGGAGGAGAACTTGGAGCCAGGTTCTCCCTGCCAAAA 4143  
Qy 4021 AGGGGCTAGATGAGTGTGCGGCGCTGGACAGCTGAGAGTGGGATTCATCCAGACTC 4080  
Db 4144 AGGGGCTAGATGAGTGTGCGGCGCTGGACAGCTGAGAGTGGGATTCATCCAGACTC 4203  
Qy 4081 ATGCAATAACCTTTGATTTGTTTCTAAAGAGGAGCTCCCTCGGCAAGATGGCAGGGT 4140  
Db 4204 ATGCAATAACCTTTGATTTGTTTCTAAAGAGGAGCTCCCTCGGCAAGATGGCAGGGT 4263  
Qy 4141 ACGGAGTCTTCAGGCCAGTTCTCTCACTTTAGCCAAATTCGAGGGCTCCTTGTGTGGAT 4200  
Db 4264 ACGGAGTCTTCAGGCCAGTTCTCTCACTTTAGCCAAATTCGAGGGCTCCTTGTGTGGAT 4323  
Qy 4201 CAGAACTAATCCAGAGTGTGGGAAAGTGACAGTCAAAACCCCACTGGAGCAAAATAAAAA 4260  
Db 4324 CAGAACTAATCCAGAGTGTGGGAAAGTGACAGTCAAAACCCCACTGGAGCAAAATAAAAA 4383  
Qy 4261 AACATACAAAGCTAATAAAAAAAAAAAAAAAAAAAAA 4293  
Db 4384 AACATACAAAGCTAATAAAAAAAAAAAAAAAAAAAAA 4416

RESULT 2  
AAZ37095  
ID AAZ37095 standard; DNA; 4192 BP.  
XX  
AC AAZ37095;  
DT 27-MAR-2000 (first entry)  
XX  
DE DNA encoding de novo DNA cytosine methyltransferase Dnmt3a.  
XX  
KW De novo DNA cytosine methyltransferase; Dnmt3a; neoplastic disorder;  
KW carcinoma; sarcoma; leukaemia; DNA methylation; ss.  
OS Mus sp.  
XX  
FH Key Location/Qualifiers  
FT CDS 217..2943  
FT /\*tag= a  
FT /product= "de novo DNA cytosine methyltransferase  
FT Dnmt3a"  
XX  
PN WO9967397-A1.  
XX  
XX 29-DEC-1999.  
XX 25-JUN-1999; 99WO-US14373.  
XX  
XX 25-JUN-1998; 98US-0090906.  
PR 24-JUL-1998; 98US-0093993.  
XX  
XX (GENO ) GEN HOSPITAL CORP.  
XX  
XX Li E, Okano M, Xie S;  
XX  
XX WPI; 2000-106298/09.  
DR P-PSDB; AAY54055.  
XX  
XX New mouse and human polypeptides, useful to treat and diagnose  
PT neoplastic disorders e.g. carcinomas, sarcomas and leukemias -  
XX  
XX Claim 8; Fig 1A; 114pp; English.  
PS

XX The present sequence encodes a murine de novo DNA cytosine  
CC methyltransferase designated Dnmt3a. The polypeptides can be administered  
CC therapeutically, especially by expressing encoding polynucleotides, to  
CC treat diseases associated with DNA cytosine methyltransferase, such as  
CC neoplastic disorders e.g. carcinomas, sarcomas and leukemias. They can  
CC be used to diagnose, or determine susceptibility to neoplastic disorders,  
CC by assaying for polypeptide expression levels in mammalian cells/body  
CC fluids. They are useful to screen for compounds inhibiting/activating the  
CC polypeptide. The polypeptides can also be used for in vitro de novo  
CC methylation of DNA. Such in vitro methylation may be used to direct or  
CC regulate DNA expression in biological systems, e.g. recombinant DNA  
CC methylation in vitro may be introduced into a cell/organism to increase  
CC or decrease expression of a desired polypeptide for which the native DNA  
CC is under-methylated or not methylated. The polypeptides can also be  
CC used to produce antibodies which are useful to detect and purify the  
CC polypeptide or therapeutically e.g. to treat neoplastic disorders. The  
CC polynucleotides are useful to produce probes and primers which are  
CC useful diagnostically.  
XX  
SQ Sequence 4192 BP; 1095 A; 1070 C; 1182 G; 844 T; 1 other;  
Query Match 66.9%; Score 2870.2; DB 21; Length 4192;  
Best Local Similarity 83.7%; Pred. No. 0;  
Matches 3573; Conservative 0; Mismatches 584; Indels 113; Gaps 24;  
Qy 24 CGGGCGGGCCCGACCCACCGGCCATACGGTGGAGCCATCGAAGCCCCCCCCACACAGGCT 83  
Db 17 CCGGGCGGGCCCGACCCCGGCCACACGGGAGCCGCCCTGAAGCCACGCGCTGAGGCT 76  
Qy 84 GACAGAGGCACTGTTACAGAGGGCTCAACACCGGGATCTATGTTTAACTTCT 143  
Db 77 G-----CACTTTCCGAGGGCTTGACATCAGGGTCTATGTTTAACTTCT 126  
Qy 144 CGGCTCCAAAGACACGATAATCTCTTCCCCCAAGCCCC-AGCAGCCCCCCCC-CCCGG 201  
Db 127 TGCTTACAAAGACACGGCAATCTCTTCTGAGGCCCTCGCAGCCCCCAGCGCCCTCG 186  
Qy 202 CAGCCCGAGCTGCTTCCGGCGCCGAGATGCCGCCATGCCCTCAGGGCCCCGGGA 261  
Db 187 CAGCCCGAGCTGCTGCGGCTACTTGCCTCA-----GCAATGCCCTCAGGGCCCCGGGA 239  
Qy 262 CACACAGCTCTCTCTCGGAGCGGGAGGAGCCGAAAGGACGAGAGGAGCAGAGGA 321  
Db 240 CACACAGCTCTCTCTCGAGCGGGAGGATGATCGAAAGGAGGAGGAGGAGGA 299  
Qy 322 GCGCGTGGCAAGGAGGAGCGCCAGAGCCAGCACACGACGACGAGGAGGTGGGCGGC 381  
Db 300 GAACCGTGGCAAGGAGGAGCGCCAGAGGCCAGCGCCACGCGCCGAGGAGGTGGGAGGCC 359  
Qy 382 TGGGAGGAGCGCAAGCAACCCCGGTGGAAAGCGGTGACACGCGCAAGGACCCCTGCGGT 441  
Db 360 TGGCGGGAAGCGCAAGCAACCCCGGTGGAAAGCAGTGACACCCCGCAAGGAGCCAGCAGT 419  
Qy 442 GATCTCCCAAGTCCCATTCATGGCCAGGACTCAGGCGCTCAGAGCTATTACCCCAATGG 501  
Db 420 GACCAACCAAGTCTCAGCCCATGGCCAGGACTCTGGCCCTCAGATCTGCTACCCATGG 479  
Qy 502 GGACTTGGAGAGCGGAGTGAAGCCCGCCAGCCAGAGAGGGAGCCCTGCTGGGGGGCAGAA 561  
Db 480 AGACTTGGAGAGCGGAGTGAAGCCCGCCAGCCAGAGAGGGAGCCAGCTGCGAGGGCAGAA 539  
Qy 562 GGGCGGGGGCCCGACGAGGAGGAGGGTGCAGCTGAGACCCCTGCTGAGGCTTCAAGAGC 621  
Db 540 GGGTGGGGCCCCAGCTGAAGGAGAGGG---AACTGAGAGCCCCACAGAGCTCCAGAGC 596  
Qy 622 AGTGGAAATGGCTGCTGACCCCGCCAGAGGAGCGGAGGAGCCCTGCGAGAGCGGCA 681  
Db 597 TGTGGAAATGGCTGCTGCTGACCAAGGAGGCGCTGAGGCTCTGAGGAGAGGGCA 656  
Qy 682 AGAACAGAGAGGAGCAACATCGAATCGAATCGAATCGAATCGAATCGAATCGAATCGA 741  
Db 657 AGAACAGAGAGGAGCAACATCGAATCGAATCGAATCGAATCGAATCGAATCGAATCGA 716



OY 742 GGGTGGCTTTGGGCTGGGAGTCCAGCTCCGTGACGGGCCCATGCGGAGGCTCACTTCCA 801  
DB 717 AGGTGGCTTGGGCTGGGAGTCCAGCTCCGTGACGGGCCCATGCGGAGGCTCACTTCCA 776  
OY 802 GGGGGGGGACCCCTACTACATCAGCAAGGCAAGCGGGAAGAGTGGTGGCAAGCTGGAA 861  
DB 777 GGGAGGGGACCCCTACTACATCAGCAAGGCAAGCGGGAAGAGTGGTGGCAAGCTGGAA 836  
OY 862 AAGGAGGCTGAGAGAGCAAGCTGATTTGAGAGATGATGTTGGAAGAAACCA 921  
DB 837 AAGGAGTCTGAGAGAGCAAGCTGATTTGAGAGATGATGTTGGAAGAAACCA 896  
OY 922 GGGGCCCCGGGAGTCTCAGAGAGTGGAGAGGCGAGCCCTCCTGTTGTCAGAGACCCAC 981  
DB 897 GGGCTTGGAGAGTCTCAGAGAGTGGAGAGGCGAGCCCTCCTGTTGTCAGAGACCCAC 956  
OY 982 TGACCCCGCATCCCGCACTGGGTACCAAGCCCTGAGCCCGTGGGGTCCGATGCTGGGA 1041  
DB 957 GGACCTCTCTTCCAGCTGGCCACCCCTGAGCCAGTATGAGGGGATGCTGGGA 1016  
OY 1042 CAAGAAATGCTACCAAGCAGGCGATGACGAGCCAGTACGAGGACGGCGGGGCTTGG 1101  
DB 1017 CAAGAAATGCTACCAAGCAGGCGATGACGAGCCAGTATGAGGGGATGCTGGGA 1076  
OY 1102 CATTGGGAGTGTGTGTGGGGAAATCTGGGGCTTCTCCTGTTGGCCAGGCGCATTTGT 1161  
DB 1077 CATTGGAGAGTGTGTGTGGGGAAATCTGGGGCTTCTCCTGTTGGCCAGGCGCATTTGT 1136  
OY 1162 GTCTTGTGATGACGGGCGGAGCGGAGCAGCTGAGAGCAGCCGCTGGGTCTGTGGTT 1221  
DB 1137 GTCTTGTGATGACGGGCGGAGCGGAGCAGCTGAGAGCAGCTGAGAGCAGCTGTGGTT 1196  
OY 1222 CGGAGACGGCAAAATCTCAGTGTGTGTGTTGAAAGCTGATGCGGCTGACTGTTTG 1281  
DB 1197 CGGAGATGCAAGTCTCAGTGTGTGTGTTGAAAGCTGATGCGGCTGACTGTTTG 1256  
OY 1282 CAGTGCCTTCCACGAGGCGCAGTCAACCAAGCAGCCATGATGCGCAAGGCTTACGA 1341  
DB 1257 CAGTGCATTTCCACGAGGCGCAGTCAACCAAGCAGCCATGATGCGCAAGGCTTACGA 1316  
OY 1342 GGTCTGAGTGTGCGCAGCAGGCGGCGGGAACTGTTCCCGTGTGCGACAGCAGA 1401  
DB 1317 AGTCTTCCAGTGTGCGCAGCAGGCGGCGGGAACTGTTCCCGTGTGCGACAGTGA 1376  
OY 1402 TGAAGTGAAGTGAAGGCGGTGAGGTGCAAGCAAGCCATGATGAGTGGCCCT 1461  
DB 1377 TGAAGTGAAGTGAAGGCGGTGAGGTGCAAGCAAGCCATGATGAGTGGCCCT 1436  
OY 1462 GGGGGGCTTCCAGCTTCTGGCCCTTAAGGGCTGAGGCCACAGAAAGAGAAATTC 1521  
DB 1437 CGGTGGCTTCCAGCCCTCGGGCTTAAGGGCTGAGGCCACAGAAAGAGAAATTC 1496  
OY 1522 CTACAAAGAGTGTACAGGAGATGTGGTGAACCTGAGGAGCTGCTTACGACACAC 1581  
DB 1497 TTACAAAGAGTGTACAGGAGATGTGGTGAACCTGAGGAGCTGCTTACGACACAC 1556  
OY 1582 TCCAACGACCAAAAGCCCGGAGAGCAGCGGAGAAAGCCCAAGTCAAGAGATTA 1641  
DB 1557 CCAACGACCAAAAGCCCGGAGAGCAGCGGAGAAAGCCCAAGTCAAGAGATTA 1616  
OY 1642 TGATGAGGCGACAGAGAGCGGTGTGTACGAGTGCAGAGAGTGCAGAAATGA 1701  
DB 1617 TGATGAGGCGACAGAGAGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1676  
OY 1702 GGACATCTGATCTCTGTGGAGGCTCAATGTTACCTTGAAACACCCCTCTTGTGG 1761  
DB 1677 GGACATCTGATCTCTGTGGAGGCTCAATGTTACCTTGAAACACCCCTCTTGTGG 1736  
OY 1762 AGGAATGTGCCAAAGCTGCAAGAACTGTTTCTGAGTGTGCGTACAGTACAGCAGA 1821  
DB 1737 AGGATGTGCCAGAACTGTAAAGAACTGTTTCTGAGTGTGCGTACAGTACAGCAGA 1796

OY 1822 CGGCTAACGCTCTACTGCAACATCTGCTGTGGGGGCCGTGAGGTCATATGCGGAA 1881  
DB 1797 TGGGTACAGCTCTATTTGACACATCTGTGGGGGGCGTGAAGTCTATGTGTGAA 1856  
OY 1882 CAACAACTGCTGAGGTGCTTTGCTGTGGAGTGTGACCTCTTGTGTGGGCGGGGGC 1941  
DB 1857 CAACAACTGCTGAGGTGCTTTGCTGTGGAGTGTGACCTCTTGTGTGGGCGGGGGC 1916  
OY 1942 TGCCAGGAGCATTAAGAGAGCCCTGGAAGTGTATATGTGCGGCAAGAGGTAC 2001  
DB 1917 TGCTCAGGAGCATTAAGAGAGCCCTGGAAGTGTATATGTGCGGCAAGAGGTAC 1976  
OY 2002 CTACGGGCTGCTGCGCGGGGAGGAGTGGCCCTCCGCGCTCAGATGTTCTGCTAA 2061  
DB 1977 CTATGGGCTGCTGCGGAGAGAGAGTGGCCCTCTCTGCACTCAGATGTTCTTCCAA 2036  
OY 2062 TAACCAAGACAGAAATTTGACCCCTCAAGGTTTACCCGACTGCCAGTGAAGAG 2121  
DB 2037 TAACCAATGACAGAAATTTGACCCCTCAAGGTTTACCCGACTGCCAGTGAAGAG 2096  
OY 2122 GAAGCCCATCCGGGCTGTCTCTTGTGATGGAATCGTACAGGAGCTCTGGTGTGAA 2181  
DB 2097 GAAGCCCATCCGGGCTGTCTCTTGTGATGGAATCGTACAGGAGCTCTGGTGTGAA 2156  
OY 2182 GGAATTTGGGCAATCAGGTGAGCCGCTACATTCCTCGAGAGTGTGTGAGATCCATCAC 2241  
DB 2157 GGAATTTGGGCAATCAGGTGAGCCGCTACATTCCTCGAGAGTGTGTGAGATCCATCAC 2216  
OY 2242 GGTGGGCAATGTCGCGCAGGAGGAGATCATGATGATGAGGAGCGTCCGAGCGCTCAC 2301  
DB 2217 GGTGGGCAATGTCGCGCAGGAGGAGATCATGATGATGAGGAGCGTCCGAGCGCTCAC 2276  
OY 2302 ACAGAGATATTCAGAGAGTGGGGCCCATTCGATCTGTGATTTGGGGAGTCCCTGCA 2361  
DB 2277 ACAGAGATATTCAGAGAGTGGGGCCCATTCGATCTGTGATTTGGGGAGTCCCTGCA 2336  
OY 2362 TGAAGTGTCAATGTCAGGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2421  
DB 2337 TGAAGTGTCAATGTCAGGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2396  
OY 2422 CTTTGAAGTGTCAATGTCAGGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2481  
DB 2397 CTTTGAAGTGTCAATGTCAGGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2456  
OY 2482 CTTTGAAGTGTCAATGTCAGGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2541  
DB 2457 CTTTGAAGTGTCAATGTCAGGAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2516  
OY 2542 ATTTCTGAGTCCAACCTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2601  
DB 2517 ATTTCTGAGTCCAACCTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2576  
OY 2602 CCGCTACTTGTGGGATTAACCTTCCCGATGATGAACAGGCGGTTGGCATCACTGTGAATGA 2661  
DB 2577 CCGCTACTTGTGGGATTAACCTTCCCGATGATGAACAGGCGGTTGGCATCACTGTGAATGA 2636  
OY 2662 TAAAGTGGAGTGCAGAGAGTGTGTGAGCATGAGCAGATGAGCAGATGAGCAGAAAGTGA 2721  
DB 2637 TAAAGTGGAGTGCAGAGAGTGTGTGAGCATGAGCAGATGAGCAGATGAGCAGAAAGTGA 2696  
OY 2722 GACCATTACTACAGAGTCAAACTTCCATTAAGCAGGAGCAAGCAGCAATTTCTGTCTT 2781  
DB 2697 GACCATTACTACAGAGTCAAACTTCCATTAAGCAGGAGCAAGCAGCAATTTCTGTCTT 2756  
OY 2782 CATGAATGAGAAAGAGAGATCTTAATGTGTGCACTGAATATGAAAGGATATTTGTGTTCC 2841  
DB 2757 CATGAATGAGAAAGAGAGATCTTAATGTGTGTGCACTGAATATGAAAGGATATTTGTGTTCC 2816  
OY 2842 AGTTCATTAAGTACAGTGTCTCAACATGAGCGCTTTGGCAGAGCAGAGATGCTGGCGG 2901  
DB 2817 AGTTCATTAAGTACAGTGTCTCAACATGAGCGCGCTTTGGCAGAGCAGAGATGCTGGCGG 2876  
OY 2902 GTATGAGGCGGTGCAAGTATCCGCCACCTTCCGCTCCGCTGAAGAGATTTTGTGCGTG 2961

Db 2877 ATCTGGAGCGTCCGGTCACTCCGACCTCTTCGCTCCGCTGAAGGAATATTTTGGTTG 2936  
Qy 2962 TGTGTAAGGACATGGGGCAAACTGAGGTAGCGA-----CACAAGTTTAAACAAACAAC 3017  
Db 2937 TGTGTAAGGACATGGGGCAAACTGAGGTAGTATGATATAAAAGTTTAAACAAACAAC 2996  
Qy 3018 -----AAAAACACAAAAACATAATAAAACACCAAGAGCATGAGGTGAGAGAGTATCA 3072  
Db 2997 AAACAAAAACAAAAACAAAAACATAAAACACCAAGAGCATGAGGTGAGAGAGT-TC 3055  
Qy 3073 GCACCCAG 3132  
Db 3086 GCACCCAG 3113  
Qy 3133 CTTTGGCTTGCAGAAAGGTTGACATCATCTCTGATTTTCAATGTTTATTTTCAGTC 3192  
Db 3114 GCTTGGCTTGCAGAAAGGTTGACATCATCTCTGATTTTCAATGTTTAACTTTCACTC 3173  
Qy 3193 CTATTTAAAAACAAACCAAGCTCCCTTCCTCTCCCTTCCTCTCTCTCTCTCTCTCTCT 3252  
Db 3174 CTATCTAAAAAGCAAAATAGGC-CCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3229  
Qy 3253 GACCTTTTATTTCTACTC-TTTTCAGAGGGTTTCTCTGTTTGTGTTGGG-TTTTGTCTT 3310  
Db 3230 AACTTTTGTCTTCTACTCTTTTTCAGAGGGTTTCTCTGTTTGTGTTGGGTTTGTCTT 3289  
Qy 3311 TGCTGTGACTGAACCAAGAGGTTATTTCAGCAAAATTCAGTAAACAAAAATAGTAAACA 3370  
Db 3290 TGCTGTGACTGAACCAAGAGGTTATTTCAGC-AAAAATCAGTAAACAAAAATAGTAA 3348  
Qy 3371 TACCTTGAG 3430  
Db 3349 TGCTTTGGAGAGAGAG-----GGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3397  
Qy 3431 TATTGGGTTGTTTTTTTGTGTTTTTGTGTTTTTGTGTTTTTGTGTTTTTGTGTTTTT 3490  
Db 3398 TATTGGTTTTTTTTTTTCTCTT-----TC 3424  
Qy 3491 TATATCTTTTTTTTGTGTTCTAGCTGATCAGATAGGAGCAAGCAGGCGGAGAGAGAG 3550  
Db 3425 TATATATCTTTTGTGTTCTTCTAGCTGATCAGATAGGAGCAAAACA--GGAAGAGAA 3482  
Qy 3551 GAGAGAGACTCAGCGGAGAGATTCCTCCAGCAGTCTGAGTGTCTGTCGACGACCA 3610  
Db 3483 TAGAGACCTCGGAGGAG-AGTCTCTCTCCAGCCCGGAGGAGTCTCAACAGCAGCA 3541  
Qy 3611 TTCCTGGTCAACGCAAAACAGAACCCAGTTAGCAGCAGGAGAGAGAGAGAGAGAGAG 3670  
Db 3542 TTCCTGGTCAACGCAAAACAGAACCCAGTCTAGCAGCAGGCGCTGAGAGAACACCA 3601  
Qy 3671 CA-TTTTCTACAGTATTTTCTAGTCTTACACAGAGAGAGAGAGAGAGAGAGAGAG 3729  
Db 3602 GACACTTTCTACAGTATTTTCTAGTCTTACACAGAGAGAGAGAGAGAGAGAGAGAG 3661  
Qy 3730 CTAGAGCGGCTTTACCTCTTGTGTTACAGTTTATATATATATATATATATATATAT 3787  
Db 3662 CTAGAGCGGCTTTACCTCTTGTGTTTACAGTTTATATATATATATATATATATAT 3721  
Qy 3788 TATATATAAAGGTACTGTTTAACTTACTGTACACCCAGTCTCATATGCTGCTTTC-AAA 3846  
Db 3722 TATATATAAAGGTACTGTTTAACTTACTGTACATCCAGTCTCATATGCTGCTTTCAAA 3781  
Qy 3847 GAGCGAGATGAGTAAACATCAGCTTCCAGTTGCTTCTGCGCAAGAGGGTTTCCACCA 3906  
Db 3782 GAGCGAGATGAGCAAAAGACATCAGTTCCGCTCGGCTCTGTGCAAAAGGGTTTCCAG 3841  
Qy 3907 GGATGAG 3966  
Db 3842 AGATGGGAG 3901  
Qy 3967 GTTTGTAAAGTGAAGGAG 4026

Db 3902 ACCCCCCAC-----CCTGCCCCATGCTAGCTTCACTTCCACCTGCAAAAAGGGG 3948  
Qy 4027 CTAGATGAGGTGTCGGGCGGCTGGACAGCTGAGAGTGGGATTTCATCCAGACTCATGCA 4086  
Db 3949 TCAGCTGAGGTGTCGGGCGGCTGGGAGAGCTGAGTGTGGAATTTATCCAGACTCGG 4008  
Qy 4087 TAACCTTTGATGTTTCTTAAAGAGAGACTCCCTCGCAAGATGCGAGAGGGTACGGAG 4146  
Db 4009 AATAACCTTTAGAAATATGAATCTAAATGACTGCTCAGAAAAATGCG-----TTGAG 4062  
Qy 4147 TCTTCAGGCGGCTTCTCACTTTAGCCAAATTCAGGGCTCCTTGTGGTGGATCAGAAC 4206  
Db 4063 AACATTTCTCTGATTTTGAATTCGTGAGCCAGCTTGAAGGCCCTTGTGGATCAGAA 4122  
Qy 4207 TAATCAGAGTGTGGGAAAGTGACAGTCAA--AACCCACCTGGAGAGCAATAAAAACA 4264  
Db 4123 TATTCAGAGTGTGGGAAAGTGACCCGCCATTAAACCCCTGGAGCAATAAAAACA 4182  
Qy 4265 TACAAAAACGT 4274  
Db 4183 TACAAAAATGT 4192

RESULT 3  
ABL90391  
ID ABL90391 standard; cDNA; 2938 BP.  
XX  
AC ABL90391;  
XX  
DT 24-MAY-2002 (first entry)  
XX  
DE Human polynucleotide SEQ ID NO 953.  
XX  
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antitaller; hepatotropic; antidiabetic; antinflammatory; antitumor;  
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW candiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein; gene; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200190304-A2.  
XX  
PD 29-NOV-2001.  
XX  
PF 18-MAY-2001; 2001WO-US16450.  
XX  
PR 19-MAY-2000; 2000US-205515P.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Birse CE, Rosen CA;  
XX  
XX WPI; 2002-122018/16.  
DR P-PSDB; ABB89982.  
XX  
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and  
PT prevention of neural, immune system, muscular, reproductive,  
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative  
PT disorders -  
XX  
PS Claim 4; SEQ ID NO 953; 2081pp + Sequence Listing; English.  
XX  
CC The invention relates to novel genes (ABL90449-ABL90853) and proteins  
CC (ABB9040-ABB9044) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune,  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's

Db 1929 GGAGAGAGGAAAAA--GGAATTCATAGAAATCTATAATATTGGGTGCTTTTTTTTTT 198

	FH	Key CDS	Location/Qualifiers
Qy	FT	CDS	115..2676 /*tag= a /product= "de novo DNA cytosine methyltransferase DNMT3B1"
Db	FT	exon	1..108 /*tag= b /number= 1 109..256 /*tag= c /number= 2 257..318 /*tag= d /number= 3 319..420 /*tag= e /number= 4 421..546 /*tag= f /number= 5 547..768 /*tag= g /number= 6 769..927 /*tag= h /number= 7 928..1035 /*tag= i /number= 8 1036..1180 /*tag= j /number= 9 1181..1240 /*tag= k /number= 10 1241..1366 /*tag= l /number= 11 1367..1411 /*tag= m /number= 12 1412..1491 /*tag= n /number= 13 1492..1604 /*tag= o /number= 14 1605..1788 /*tag= p /number= 15 1789..1873 /*tag= q /number= 16 1874..2019 /*tag= r /number= 17 2020..2110 /*tag= s /number= 18 2111..2259 /*tag= t /number= 19 2260..2345 /*tag= u /number= 20 2346..2415 /*tag= v /number= 21 2416..2534 /*tag= w /number= 22 2535..4145 /*tag= x
Qy	FT	exon	3451 TGTTTTTG-TTTTTTTTGGGTTTTTTTTTTTTTACTATATACTTTTTTTGTGT 3509
Db	FT	exon	1986 TWTNTTNTTTTTTTTTTTGGGTTTTTTTTTTTTTACTATATACTTTTTTTGTGT 2045
Qy	FT	exon	3510 CTCTAGCCTCATCAGATAGGACCAACAGCGGGACGGAAGAGAGAGACACTCAGCGG 3569
Db	FT	exon	2046 CTCCTAGCCTCATCAGATAGGACCAACAGCGGGACGGAAGAGAGAGACACTCAGCGG 2105
Qy	FT	exon	3570 CAGCATTCCTCCAGCCACTGAGCTGTGTCGCCAGCACATTCTCTGGTCCAGCAAACA 3629
Db	FT	exon	2106 CACATTTCCCTCCAGCCACTGAGCTGTGTCGCCAGCACATTCTCTGGTCCAGCAAACA 2165
Qy	FT	exon	3630 GAACCCAGTTAGCAGCAGGAGACGAGAACACCACACAGACATTTTCTACAGTATTTC 3689
Db	FT	exon	2166 GAACCCAGTTAGCAGCAGGAGACGAGAACACCACACAGACATTTTCTACAGTATTTC 2225
Qy	FT	exon	3690 AGGTGCCTACACACAGAAACCTTTGAAGAAAATCAGTTTCTAGAAGCCGCTGTACCTC 3749
Db	FT	exon	2226 AGGTGCCTACACACAGAAACCTTTGAAGAAAATCAGTTTCTAGAAGCCGCTGTACCTC 2285
Qy	FT	exon	3750 TTGTTTACAGTTTATATATATATAGATATAGATATATATAAAGGTACTGTTAA 3809
Db	FT	exon	2286 TTGTTTACAGTTTATATATATATAGATATAGATATATATAAAGGTACTGTTAA 2345
Qy	FT	exon	3810 CTACTGFPAACCCGACTTCATAATATGGTGCTTTCAACACGAGAGATGATTAACATCA 3869
Db	FT	exon	2346 CTACTGFPAACCCGACTTCATAATATGGTGCTTTCAACACGAGAGATGATTAACATCA 2405
Qy	FT	exon	3870 GCTTCCACGTTGCCCTCTGCGCAAAAGGTTTTCACCAAGATGAGAAAGGAGACAGCTT 3929
Db	FT	exon	2406 GCTTCCACGTTGCCCTCTGCGCAAAAGGTTTTCACCAAGATGAGAAAGGAGACAGCTT 2465
Qy	FT	exon	3930 GCAGATGGCGGTTCTCACGGTGGGCTCTTCCCCTTTGGTTGTGTAACGAAAGTGAAGGAGGA 3989
Db	FT	exon	2466 GCAGATGGCGGTTCTCACGGTGGGCTCTTCCCCTTTGGTTGTGTAACGAAAGTGAAGGAGGA 2525
Qy	FT	exon	3990 GAACTTTGGGAGCCAGGTTCTCCCTGCCAAAAAGGGGCTTAGATGAGGTGGTTCGGGCCCGT 4049
Db	FT	exon	2526 GAACTTTGGGAGCCAGGTTCTCCCTGCCAAAAAGGGGCTTAGATGAGGTGGTTCGGGCCCGT 2585
Qy	FT	exon	4050 GGACAGCTGAGTGGGATTCATCCAGACTCATGCAATACCCCTTGATTTGTTTCTTAAA 4109
Db	FT	exon	2586 GGACAGCTGAGTGGGATTCATCCAGACTCATGCAATACCCCTTGATTTGTTTCTTAAA 2645
Qy	FT	exon	4110 AGGAGACTCCCTCCGGCAAGATGGCAGAGGATACGGAGTCTTCAGGCCAGTTTCTCACTT 4169
Db	FT	exon	2646 AGGAGACTCCCTCCGGCAAGATGGCAGAGGATACGGAGTCTTCAGGCCAGTTTCTCACTT 2705
Qy	FT	exon	4170 TAGCCAATTGAGGGCTCTTGTGTGGATCAGAACTAATCCAGACTGTCGGAAGTGA 4229
Db	FT	exon	2706 TAGCCAATTGAGGGCTCTTGTGTGGATCAGAACTAATCCAGACTGTCGGAAGTGA 2765
Qy	FT	exon	4230 CAGTCAAAACCCACCTCGAGCAATATAAAAAACATACAAAAAGTA 4275
Db	FT	exon	2766 CAGTCAAAACCCACCTCGAGCAATATAAAAAACATACAAAAAGTA 2811
RESULT 4			
AZ37098	ID		AAZ37098 standard; DNA; 4145 BP.
XX	AC		AAZ37098;
XX	DT		27-MAR-2000 (first entry)
XX	DE		DNA encoding de novo DNA cytosine methyltransferase DNMT3B1.
XX	KW		De novo DNA cytosine methyltransferase; DNMT3B1; neoplastic disorder;
XX	KW		carcinoma; sarcoma; leukaemia; DNA methylation; ss.
XX	OS		Homo sapiens.
XX	OS		



Db 2080 AATGTGAATCCAGCCAGGAAGGCTGTATGAGGGTACAGCGCGCTCTTCTCGAATTT 2139  
Qy 2432 TACCGCTCTCTGATGATGGCGGCCCAAGGAGGAGATGATCGCCCTTCTTCTGGCTC 2491  
Db 2140 TACCACCTCTGTAATCTACACGCCCCCAAGGAGGGTGTATGACCGCGCTTCTTCTGGATG 2199  
Qy 2492 TTGTGAGATGTGTGGCCATGTGGCGCTTAGTGAACAAGAGGACATCTCCGATTTCTCGAG 2551  
Db 2200 TTTGAGATGTTGTAGCCATGAAGTTGGCGACAAGAGGACATCTACCGTTCTCGAG 2259  
Qy 2552 TCCAACTCTGTATGATGTCGAAGAAGTGTACGTGCACACAGGCCCGCTACTTC 2611  
Db 2260 TGTATCCAGTATGATGATGTCATCAAAAGTTTCTGTCTACAGGGCCCGATATTC 2319  
Qy 2612 TGGGCTAACCTTCCCGGTATGAACAGGCCGTTGGCATCCACTGTGAATGATAAGCTGGAG 2671  
Db 2320 TGGGCAACCTACCCCGGATGAACAGGCCCGTGTATGATCAAAAGATGATAAACTCGAG 2379  
Qy 2672 CTGAGGAGTGTCTGAGCATGCGAGATAGCCAAAGTTCAGCAAGTGAGGACCAATTACT 2731  
Db 2380 CTGAGGAGTGTCTGGAATACAAATAGGATAGCCAAAGTTAAAGAAGTACAGCAATAACC 2439  
Qy 2732 ACGAGTCAAACTCCATAAAGCAGGGCAAGACACAGCATTTTCTGTCTTATGAATGAG 2791  
Db 2440 ACCAAGTCAACTCGATCAAAACAGGGGAAAAACCAACTTTTCCCTGTGTGATGAATGCG 2499  
Qy 2792 AAGAGGACATCTTATGCTGCTCACTGAATGGAAGGGTATTTGGTTTCCAGTCCACTAT 2851  
Db 2500 AAGAGGATGTTTGTGTGCTGCTGAGCTGGAAGGATCTTTGGCTTTCTGTGCACTAC 2559  
Qy 2852 ACTGAGTCTCCAACTAGCCGCTTGGCGAGGAGAGATCTGCTGGCGGCTCATGGAGC 2911  
Db 2560 ACAGAGTGTCCAACTAGGCGCGTGTGTCGCCGCCAGAGCTGCTGGGAAGGTCTCTGGAGC 2619  
Qy 2912 GTGCAGTATCATCGCCACCTCTTCTGCTCCGCTGAAGGAGTATTTTGGCTGTGTGTA 2967  
Db 2620 GTGCTGTATCGACACCTCTTCTGCCCCCTCTGAAGGACTACTTTGCAATGTGAATA 2675

RESULT 5  
AAZ37096  
ID AAZ37096 standard; DNA; 4195 BP.  
XX  
AC AAZ37096;  
XX  
DT 27-MAR-2000 (first entry)  
XX  
DE DNA encoding de novo DNA cytosine methyltransferase Dnmt3b1.  
XX  
KW De novo DNA cytosine methyltransferase; Dnmt3b1; neoplastic disorder;  
KW carcinoma; sarcoma; leukaemia; DNA methylation; ss.  
XX  
OS Mus sp.  
XX  
PH Key Location/Qualifiers  
FT CDS 269..2848  
FT /tag= a  
FT /product= "de novo DNA cytosine methyltransferase  
Dnmt3b1"  
XX  
FN WO967397-A1.  
XX  
PD 29-DEC-1999.  
XX  
PF 25-JUN-1999; 99NO-US14373.  
XX  
PR 25-JUN-1998; 98US-0090906.  
PR 24-JUL-1998; 98US-0093993.  
XX  
PA (GEO ) GEN HOSPITAL CORP.  
XX  
PI Li E, Okano M, Xie S;

XX WPI: 2000-106298/09.  
DR P-PSDB; AAY54056.  
XX New mouse and human polypeptides, useful to treat and diagnose  
PT neoplastic disorders e.g. carcinomas, sarcomas and leukemias -  
XX Claim 9; Fig 1B; 114pp; English.  
XX The present sequence encodes a murine de novo DNA cytosine  
CC methyltransferase designated Dnmt3b1. The Dnmt3b gene also produces,  
CC through alternate splicing, at least two shorter isoforms of 840 and  
CC 777 amino acid residues, termed Dnmt3b2 and Dnmt3b3 (sequences  
CC not given in the specification). The polypeptides can be administered  
CC therapeutically, especially by expressing encoding polynucleotides, to  
CC treat diseases associated with DNA cytosine methyltransferase, such as  
CC neoplastic disorders e.g. carcinomas, sarcomas and leukemias. They can  
CC be used to diagnose, or determine susceptibility to neoplastic disorders,  
CC by assaying for polypeptide expression levels in mammalian cells/body  
CC fluids. They are useful to screen for compounds inhibiting/activating the  
CC polypeptide. The polypeptides can also be used for in vitro de novo  
CC methylation of DNA. Such in vitro methylation may be used to direct or  
CC regulate DNA expression in biological systems, e.g. recombinant DNA  
CC methylated in vitro may be introduced into a cell/organism to increase  
CC or decrease expression of a desired polypeptide for which the native DNA  
CC is under-methylated or not methylated. The polypeptides can also be  
CC used to produce antibodies which are useful to detect and purify the  
CC polypeptide or therapeutically e.g. to treat neoplastic disorders. The  
CC polynucleotides are useful to produce probes and primers which are  
CC useful diagnostically.

XX Sequence 4195 BP; 1061 A; 1059 C; 1123 G; 952 T; 0 other;  
Qy Query Match 17.7%; Score 760.4; DB 21; Length 4195;  
Best Local Similarity 63.7%; Pred. No. 6.8e-145;  
Matches 1270; Conservative 0; Mismatches 661; Indels 63; Gaps 5;  
Qy 1016 GAGCCCTGGGGTCCGATGCTGGGACAAAGATGCACCAAGAGCGGATGACGAGCCA 1075  
Db 875 GATCAGGAGGTATGGATACACACAGGTGGATGACAGAGCATATATGGAGACAGCACA 934  
Qy 1076 GAGTACAGGACGCGCGGGCTTTGGCATTTGGGAGCTGGTGTGGGGGAAACTGCGGGGC 1135  
Db 935 GAGTATCAGGATGATAAAGAGTTTGGAAATAGGTGACCTCTGTGGGGGAAAGATCAAGGC 994  
Qy 1136 TTCTCTGGTGGCCAGCGCATTTGTCTTGGTGGATGACGGCGCGAGCCGAGCAGCT 1195  
Db 995 TTCTCTGGTGGCTGCCATGCTGGTGTCTGGAAAGCCACCTCCAAAGCGACAGGCCATG 1054  
Qy 1196 GAAGGACCGCGCTGGTCACTGTTCGGAGACGGCAAAATCTCAGTGGTGTGTGTGAG 1255  
Db 1055 CCGGAATGCGCTGGTACAGTGGTTGGTGTGATGGCAAGTTTCTGAGATCTCTGCTGAC 1114  
Qy 1256 AAGCTGATGCGCTGAGCTCGTTTTCAGTGGTTCACAGGCCACGTACAAACAGCAG 1315  
Db 1115 AAATGTTGGCTCTGGGGCTGTTCAGCCAGCATTTAATCTGGCTACCTTCAATAAGCTG 1174  
Qy 1316 CCATGTACCGCAAGCCATCTACAGGTCTTCGAGGTGGCCAGCAGCCGCGGGGAAG 1375  
Db 1175 GTTTCATTATAGGAAGGCCATGTACCACACTCTGGGAAGAACCGAGGTTCAGCTGGCAAG 1234  
Qy 1376 CTGTTCCCGGTGTGCCACAGCAGCGATGAGATGACACTGCCAAGGCCGTGGAGGTGCAG 1435  
Db 1235 ACCTTCTCCAGCATC-----CTGGAGAGTCACTGGAGGACCAAG 1273  
Qy 1436 AACAAAGCCCATGATTGAATGGGCCCTGGGGGCTTCCAGCCTTCTGGCCCTTAAGGGCCTG 1495  
Db 1274 CTGAAAGCCCATGCTGGAGTGGGCCCAACCGTGGCTTCAAGCCTACTCTGGGATCGAGGGCTC 1333  
Qy 1496 GAGCC--ACCAAGAAGAAGAAATCCCTACAAAGAAAGTGTACACGGAATGTGGGTGG 1553  
Db 1334 AAACCAACAAAGAGCAACCAAGTGGTTAATAGTGAAGGTGCTGCTTCAGACAGTAGG 1393



QY 1554 AACCT-----GAGGAGCTGCTACGACACCTCCACC 1587  
Db 1394 AACTTAGAACCGAGAGACCGAGAACAAAGTCGAAGACGACCAACATGACTCTGCT 1453  
QY 1588 AGCCAAAAGCCCCGGAAGAGACACGCGAGAAAGCCCAAGGTCAAGAGATTA----- 1640  
Db 1454 GCTTCTAGTCCCCCCCCACCCCAAGCGCTCAAGACAAATAGCTATGCGGGAAGACCGA 1513  
QY 1641 -----TTGATGACGCGACAAAGAGACGCGCTGTGTAGAGGTGGCGGACAGAGTCCGGAC 1696  
Db 1514 GGGAGAGATGAGAGAGCGGAGACGATGGCTTCTGAAGTCACCAACAAAGGGAAT 1573  
QY 1697 ATTGAGGACATCTGATCTCTCTGTGGAGCTCAATGTACCTTGAACACCCCTCTTC 1756  
Db 1574 CTGGAAGACCGCTGTCTGTCTGTGGAAGAAAGAACCTGTGTCTTCAACCCCTCTTT 1633  
QY 1757 GTTGAAGAAATGTGCAAAACCTGCAAGAACTGCTTCTGAGTGTGCGTACAGTACGAC 1816  
Db 1634 GAGGTTGGGCTCTGTAGAGTGTCCGGATCGCTTCTTACAGCTCTTCAATGATATGAT 1693  
QY 1817 GACGACGCTACAGTCTCTCTCTGACCATCTGCTGTGGGCGGAGAGTGTCTATGTC 1876  
Db 1694 GAGGACGCTACAGTCTCTCTCTGACCATCTGCTGTGGGCGGAGAGTGTCTATGTC 1753  
QY 1877 GGAACAACAACACTGCTGACGAGTCTTTGCGTGAAGTGTGACCTCTTGTGGGCGG 1936  
Db 1754 AGTACACAAAGTCTGCAAGTCTTCTGTGAGTGTCTGAGAGTGTCTGTGGGCGGCA 1813  
QY 1937 GGGGCTGCCAGGACCATTAAGAAAGACCCCTGAACTGCTACATGTGCGGACAAAG 1996  
Db 1814 GGCACAGTGAAGATGCCAAGCTGCAAGAACCTGAGCTGTATGTGCTCTCCAG 1873  
QY 1997 GGTACCTACGCGCTGTGCGGCGGAGAGACTGCGCTCCGAGTCCAGATGTTCTTC 2056  
Db 1874 CGCTGCCATGGGCTCTCCAGACGAGAAAGATTGAAACATGCGCTGCAAGCTTCTTC 1933  
QY 2057 GCTAATACCAAGAC--AGAAATTGACCCCTCAAGGTTTACCACTGTCTCCAGCT 2113  
Db 1934 ACTACTGATCTGACCTGGAAGATTGAGCCACCAAGTTGACCACTTCTGCA 1993  
QY 2114 GAGAAAGAAAGCCATCCGAGTGTCTCTTTGATGGAATCGTACAGGCTCTCTG 2173  
Db 1994 GCCAAAGAGAGGCCATTAGAGTCTGTCTCTTTGATGGAATGCAACGGGCTCTTG 2053  
QY 2174 GTGCTGAAGACTTGGGCAATTCAGTGAACCGTACATTTGCTCGAGTGTGAGAC 2233  
Db 2054 GTGCTCAAGAGTGGGTATTAAGTGAAGAAAGTACCTTGTCTCGAAGTGTGTGAGAG 2113  
QY 2234 TCCATCAAGTGGGACGTGTGCGGACCAAGGAAATCATGTCGTGCGGACGTCCG 2293  
Db 2114 TCCATCGCTGTGGAACTGTTAAGCATGAAGGCAAGTCAATATGTCAATGACGTCCG 2173  
QY 2294 AGCGTCACAGAGCATATTCAGAGTGGGCGCAATTCATCTGTGATTTGGGCGAGT 2353  
Db 2174 AAATACCAAAATAATTAAGAGTGGGCGGCTTCACTTGTGATTTGTTGGAAGC 2233  
QY 2354 CCTGCAATGACCTCTCATGTCGCAACCTGTGCAAGGCGCTTCAAGGCGCATGCG 2413  
Db 2234 CCAATGCAATGATCTCTTAAGTCAATCTGCCGCAAGTTTATATAGGCGACAGGA 2293  
QY 2414 CGGCTCTTCTTGAAGTCTAACCGCTCTGATGATGCGGCGGCGCAAGGAGATGAT 2473  
Db 2294 AGGCTCTTCTTGAAGTCTTACCACTTGTGCAATTAATCCGCGCCCAAGAGGCGCAAC 2353  
QY 2474 CGCCCTTCTTCTGCTTTGAGAAATGTGTGGCCATGCGGCTTGTAGTACAGAGGAG 2533  
Db 2354 GGTTCATTTCTTCTGATGTTGAGAAATGTGTGGCCATGAAGTAAACAAAGAAAGAC 2413  
QY 2534 ATCTGCGATTTCTGAGTCCAAACCTGTGATGATGATGCGCAAGAAAGTGTGAGTCA 2593  
Db 2414 ATCTCAAGATTTCTGAGATGAACCAAGTATGATGATGATGATGATGATGATGATG 2473  
QY 2594 CACAGGCGCGCTACTTCTTGGGGTAACTTCCGGTATGAACAGGCGTTGGCATCACT 2653

Db 2474 CACAGGCGCGCTACTTCTTGGGGTAACTTCCGGATGAACAGGCGGCTGATGCTTCA 2533  
QY 2654 GTGATGATTAAGCTGAGCTGACAGAGTGTGTGAGCATGAGATGATGATGATGATGAT 2713  
Db 2534 AAGATGATTAAGCTGAGCTGACAGAGTGTGTGAGCATGAGATGATGATGATGATGAT 2593  
QY 2714 AAGTGAAGACCATTAAGCTGAGCTGACAGAGTGTGTGAGCATGAGATGATGATGAT 2773  
Db 2594 AAGTGAAGACCATTAAGCTGAGCTGACAGAGTGTGTGAGCATGAGATGATGATGAT 2653  
QY 2774 CCTGCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2833  
Db 2654 CCTGCTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2713  
QY 2834 GGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2893  
Db 2714 GGTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2773  
QY 2894 CTGGGCGGCTCATGAGCGCTGCGAGTCAATCCGCACTCTTGTGCTGCTGCTGCTG 2953  
Db 2774 CTGGGCGGCTCATGAGCGCTGCGAGTCAATCCGCACTCTTGTGCTGCTGCTGCTG 2833  
QY 2954 TTTCGCTGTGCTA 2967  
Db 2834 TTTCGCTGTGATTA 2847  
  
RESULT 6  
ABQ57540  
ID ABQ57540 standard; cDNA; 709 BP.  
XX AC ABQ57540;  
XX DT 02-AUG-2002 (first entry)  
XX DE Human colon cancer related nucleotide sequence SEQ ID NO:1235.  
XX KW Human; colon cancer; cancer; tissue profiling; forensic; mapping;  
XX KW genetic analysis; diagnostic; antisense therapy; gene; ss.  
XX OS Homo sapiens.  
XX PN WO200229086-A2.  
XX PD 11-APR-2002.  
XX PF 02-OCT-2001; 2001WO-US30732.  
XX PR 02-OCT-2000; 2000US-237271P.  
XX PA (FARB ) BAYER CORP.  
XX PI Burgess C, Astle JH, Carroll E, Catino TJ, Dwivedi P, Molino GA;  
XX PI Thiglingam A, Lewis ME;  
XX DR WPI; 2002-426115/45.  
XX PT New isolated nucleic acid that is differentially expressed in cancer  
XX PT tissues useful for determining the presence of colon cancer in a cell  
XX PT or tissue type, and in antisense therapy  
XX PS Claim 1; Fig 1; 796dp; English.  
XX  
XX ABQ56306 to ABQ60787 represent isolated nucleic acids (I) differentially  
XX expressed in cancer tissues. ABB78993 to ABB79004 represent proteins  
XX encoded by the ABQ60776 to ABQ60787 nucleic acid sequences. (I) can be  
XX used in antisense therapy. An antibody immunoreactive with a polypeptide  
XX encoded by (I) is useful for detecting cancer in a patient sample, and  
XX for detecting the presence or absence of a polynucleotide encoded by a  
XX nucleic acid which hybridizes to (I) in a cell. A probe/primer derived  
XX from (I) can be used for determining the presence of a nucleic acid which  
XX hybridizes to (I), and for determining the phenotype of cells in a sample



CC of cells from a patient. (I) is useful for determining the presence of  
CC colon cancer in a cell or tissue type, for determining the presence or  
CC state of other type of cancer, in antisense therapy, to generate  
CC macroarrays on a solid surface, to identify a chromosome on which the  
CC corresponding gene resides, and in tissue profiling, forensics, genetic  
CC analysis, mapping and diagnostic applications. (I) can be used to raise  
CC antibodies, and to screen for peptide analogues and antagonists.  
XX  
SQ Sequence 709 BP; 163 A; 177 C; 193 G; 168 T; 8 other;  
Query Match 14.3%; Score 613; DB 24; Length 709;  
Best Local Similarity 94.5%; Pred. No. 3.6e-115;  
Matches 664; Conservative 0; Mismatches 33; Indels 6; Gaps 3;  
QY 2275 GTACGTGGGGAGCTCCGAGCGTCACACAGAACATATCAGGAGTGGGCCCATTCGA 2334  
Db 1 GCACCTCGGGAGCTCCGAGCGTCACACAGAACATATCAGGAGTGGGCCCATTCGA 60  
QY 2335 TCTGTGATGGGGAGCTCCGAGCGTCACACAGAACATATCAGGAGTGGGCCCATTCGA 2394  
Db 61 TCTGTGATGGGGAGCTCCGAGCGTCACACAGAACATATCAGGAGTGGGCCCATTCGA 120  
QY 2395 CCTTACGAGGCGACTGGCGGCTCTTTTGTAGTTCTACCGCTCTCTGCATGATGCGG 2454  
Db 121 CCTTACGAGGCGACTGGCGGCTCTTTTGTAGTTCTACCGCTCTCTGCATGATGCGG 180  
QY 2455 GCCAAGGAGGAGATGATCGCCCTCTTCTGCTCTTTGAGATGTTGGTGGCCATGG 2514  
Db 181 GCCAAGGAGGAGATGATCGCCCTCTTCTGCTCTTTGAGATGTTGGTGGCCATGG 240  
QY 2515 GCTTAGTGACAGAGGAGCATCTCGGATTTCTGAGTTCACCAACCTGTGATGATGTC 2574  
Db 241 GCTTAGTGACAGAGGAGCATCTCGGATTTCTGAGTTCACCAACCTGTGATGATGTC 300  
QY 2575 CAAGAAGTGTGAGTGCACACAGGGCCCTACTTCTGGGGTAACTTCCCGGTATGAA 2634  
Db 301 CAAGAAGTGTGAGTGCACACAGGGCCCTACTTCTGGGGTAACTTCCCGGTATGAA 360  
QY 2635 CAGGCGTGGATCCATCTGTGAATAGCTGGAGCTGAGAGTGTCTGGAGCATGG 2694  
Db 361 CAGGCGTGGATCCATCTGTGAATAGCTGGAGCTGAGAGTGTCTGGAGCATGG 420  
QY 2695 CAGGATAGCAAGTTCAGCAAGTGGAGGACCATTAACGAGGTCAAACTCCATAAGCA 2754  
Db 421 CAGGATAGCAAGTTCAGCAAGTGGAGGACCATTAACGAGGTCAAACTCCATAAGCA 480  
QY 2755 GGGCAAGACCGACATTTCTCTCTTCATGATGAGAAAGGAGCATCTTATGTC-GCA 2813  
Db 481 GGGCAAGACCGACATTTCTCTCTTCATGATGAGAAAGGAGCATCTTATGTCGCA 540  
QY 2814 CTGAATGGAAGGGTATTGGTTTCCAGTCCACTATCTGAGTCTCCCAACATGAGCC 2873  
Db 541 CTGAATGGAAGGGTATTGGTTTCCAGTCCACTATCTGAGTCTCCCAACATGAGCC 600  
QY 2874 GCTTGGCGAGGAGACTGCTGGCGCGTCA-TGGAGCGTGCCA---GTCATCCGCCA 2928  
Db 601 GCTTGGCGANGCAAAACTGCTGGGCCCTCATTTGGACCGTGCCTCAATTCACCT 660  
QY 2929 CCTTTGCTCGCTGAAGAGATTTTGGTGTGTGTAGGG 2971  
Db 661 TTTTNGNTTCCGTTGAAGGAATTTTGTGTTGTTAAG 703  
RESULT 7  
AAH33860  
ID AAH33860 standard; cDNA; 1060 BP.  
XX  
AC AAH33860;  
XX  
DT 03-SEP-2001 (first entry)  
XX  
DE Human colon cancer antigen encoding cDNA SEQ ID NO:916.  
XX

KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW colorectal carcinoma; ss.  
XX  
OS Homo sapiens.  
XX  
PN W0200122920-A2.  
XX  
PD 05-APR-2001.  
XX  
PF 28-SEP-2000; 2000WO-US26524.  
XX  
PR 29-SEP-1999; 99US-0157137.  
PR 03-NOV-1999; 99US-0163280.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Ruben SM, Barash SC, Birse CE, Rosen CA;  
XX  
XX WPI: 2001-235357/24.  
XX  
XX P-PSDB; AAG74429.  
XX  
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
XX useful for preventing, diagnosing and/or treating colorectal cancers -  
XX  
XX Claim 1; Page 2829; 9803pp; English.  
XX  
XX AAH32943 to AAH37195 and AAG77788 represent human colon  
XX cancer-associated nucleic acid molecules (N) and proteins (P), where  
XX the proteins are collectively known as colon cancer antigens. The colon  
XX cancer antigens have cytostatic activity and can be used in gene  
XX therapy and vaccine production. N and P may be used in the prevention,  
XX diagnosis and treatment of diseases associated with inappropriate P  
XX expression. For example, N and P may be used to treat disorders  
XX associated with decreased expression by rectifying mutations or deletions  
XX in a patient's genome that affect the activity of P by expressing  
XX inactive proteins or to supplement the patient's own production of P.  
XX Additionally, N may be used to produce the colon cancer-associated P,  
XX by inserting the nucleic acids into a host cell and culturing the cell  
XX to express the proteins. N and P can be used in the prevention, diagnosis  
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
XX and AAB77789 represent sequences used in the exemplification of the  
XX present invention.  
XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
XX missing at time of publication, meaning no sequences are present for  
XX SEQ ID NO:1027 to 1052, 7921 and 7922.  
XX  
SQ Sequence 1060 BP; 267 A; 262 C; 297 G; 227 T; 7 other;  
Query Match 12.8%; Score 548.2; DB 22; Length 1060;  
Best Local Similarity 72.6%; Pred. No. 6.7e-102;  
Matches 700; Conservative 4; Mismatches 260; Indels 0; Gaps 0;  
QY 2004 ACGGGCTGCTGGCGGCGAGAGGACTGGCCCTCCCGCTCCAGATGTTCTTCGCTAATA 2063  
Db 18 ATGCGCTCTGCGGCGCGGAGGAGGACTGGAACGTGCGCTTCAGGCTTCTTCACCATG 77  
QY 2064 ACCACACACAGGATTTGACCCCTCAAGGTTTACCCACCTGTCAGCTGAGAGAGGA 2123  
Db 78 ACAGGGGCTTGAATACGAAGCCCCCAAGCTGTACCTCGCATCCCGAGCCGAGGCG 137  
QY 2124 AGCCCATCCGGTCTCTCTCTTTGATGAATCGTACAGGGCTCTCGTGTGCTGAAGG 2183  
Db 138 GGCCCATTCGAGTCTCTCATTTGTTGATGTCATCGCACAGGCTACCTAGTCTCTCAAG 197  
QY 2184 ACTTGGGCATTCAGGTGGACCGCTACATTCCTCGGAGGTGTGAGGACTCCATCACCG 2243  
Db 198 AGTTGGGCATTAAGGTAGGAAAGTACGTCGCTTCTGAAGTGTGTAGGAGTCCATTCCTG 257  
QY 2244 TGGGCATGTTGGGCACACAGGGGAGATCATGTACGTGGGAGCGTCCGACGCTCACAC 2303  
Db 258 TTGGAACCGTGAAGCACAGGGGGAATATCAAAATAGCTGACGAYGTGAGGAACATCACAA 317  
QY 2304 AGAAGCATATCCAGGAGTGGGCCCTTCGATCTGTTGGGGGCGAGTCCCTCGCAATG 2363







Db 1440 TGTGCACTACA 1450

RESULT 10  
AAS86165  
ID AAS86165 standard; cDNA; 711 BP.  
XX  
AC AAS86165;  
XX  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #21969.  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
XX  
XX W0200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001WO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
XX  
XX 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI: 2001-639362/73.  
XX  
XX P-PSDB; ABG21978.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX Claim 1: SEQ ID No 21969; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 711 BP; 181 A; 164 C; 199 G; 167 T; 0 other;  
XX  
XX Query Match 8.0%; Score 344.2; DB 23; Length 711;  
XX Best Local Similarity 75.0%; Pred. No. 2e-60;  
XX Matches 430; Conservative 0; Mismatches 143; Indels 0; Gaps 0;  
XX  
QY 2135 GTGCTGCTCTTTGATGGAATCGTACAGGCTCTGCTGCTGAAGGACTTGGGCATT 2194  
Db 1 GTCTGTCTATTTTGTGTCATCGGACAGGCTACTAGTCTCAAGAGTTGGGCATA 60

QY 2195 CAGGTGGACCGCTACATTGCTCGAGGTGTGTAGGACTCCATCACGGTGGGCATGTTG 2254  
Db 61 AAGGTAGGAAAGTACGTGCTTCTGAAGTGTGTAGGAGTCCATTGCTGTTGGAACCGTG 120  
QY 2255 CGCACCCAGGGGAAGATCATGTAGTCGGGACGTCGCGACGTCACAGAGCATATC 2314  
Db 121 AAGCACGAGGGGAATATCAATACGTCAACGACGTGAGGAACATCACAAAGAAAAATTT 180  
QY 2315 CAGGAGTGGGGCCCATTCGATCTGGTGAATTGGGGGCGATCCCTGCAATGACCTCCATC 2374  
Db 181 GAAGAATGGGGCCCATTTGACTTGGTGAATTGGCGGAAGCCCATGCAACGATCTCAAAAT 240  
QY 2375 GTCAACCTGCTCGCAAGGCCCTTACGAGGGCACTGGCCGGCTCTTTTGAAGTTCTAC 2434  
Db 241 GTGAATCCAGCCAGGAAGGCCCTGTATGAGGGTACAGGCCGGCTCTTCTTCCAAATTTTAC 300  
QY 2435 CGCTCTCTGATGATGCGCGGCCCAAGGAGGAGATGATCGCCCTTCTTCTGGCTCTTTT 2494  
Db 301 CACCTGCTGATTTACTCACGCCCAAGGAGGGGTATGACCGGCCGTTCTTCTGGATGTTT 360  
QY 2495 GAGATGTGGTGGCCATGGCGGTTAGTGACAAGAGGGACATCTCCGGAATTTCTCGAGTCC 2554  
Db 361 GAGATGTTGTAGCCATGAAGGTTGGCGCAAGAGGGACATCTCACGGTTCCTGGAGTGT 420  
QY 2555 AACCTGTGATGATGATGATGCAAGAAAGTGTGAGTGCACACAGGGCCCGTACTTCTCG 2614  
Db 421 AATCCAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
QY 2615 GGTAACCTTCCCGGTATGAAACAGGCCGTTGGCATCCACTGTGAATGATGAAGTGGAGCTG 2674  
Db 481 GGCACCTTACCCGGGATGAAACAGGCCGTTGATAGCATCAAGAAGATGATAAATCGAGCTG 540  
QY 2675 CAGGAGTGTCTGGAGCATGGCAGGATAGCCAAAG 2707  
Db 541 CAGGACTGCTTGGCAATACATAGGATAGCCAAAG 573

RESULT 11  
AAS80537  
ID AAS80537 standard; cDNA; 2077 BP.  
XX  
XX AAS80537;  
XX  
DT 08-JUN-2001 (first entry)  
XX  
DE Receptor #25 partial coding sequence.  
XX  
KW Probe; microarray; cancer; immunopathology; neuropathology; ss.  
XX  
XX Balaena mysticetus.  
XX  
XX US6183968-B1.  
XX  
PD 06-FEB-2001.  
XX  
XX 25-MAR-1999; 99US-0276531.  
XX  
XX 27-MAR-1998; 98US-0079677.  
XX  
XX (INCY-) INCYTE PHARM INC.  
XX  
XX Bandman O, Lal P, Hillman JL, Yue H, Reddy R, Guegler KJ;  
PI Baughn MR;  
XX  
XX WPI; 2001-201999/20.  
XX  
XX Composition having probes which comprise part of gene sequence encoding  
PT proteins associated with cell proliferation useful as hybridizable  
PT array elements in Microarrays to monitor expression of target  
XX polynucleotide -  
XX  
PS Claim 1; Columns 93-96; 104pp; English.  
XX

The present invention relates to a composition comprising several polynucleotide probes. Probes can be derived from the present sequence. The probes are immobilised and are preferably useful as hybridisable array elements in a microarray for monitoring the expression of several polynucleotides. The microarray can be used in the diagnosis of cancers such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma and tetracarcinoma etc., immunopathology such as AIDS, Addison's disease, adult respiratory distress syndrome, allergies, anaemia, asthma, atherosclerosis and bronchitis etc., neuropathology such as Alzheimer's disease, amnesia, amyotrophic lateral sclerosis, bipolar disorder, catatonia and cerebral neoplasms etc. The microarray can also be used to investigate an individual's predisposition to a disease such as cancer, immunopathology or neuropathology. Also, the microarray can be used for investigating cellular response to infection, drug treatment etc. The microarray can be used for diagnostics, prognostics and treatment of regimens, drug discovery and development, toxicological and carcinogenicity studies, forensics, pharmacogenomics etc. The array can also be used for monitoring disease progression.

Sequence 2077 BP; 539 A; 485 C; 477 G; 567 T; 9 other;

Query Match 7.6%; Score 324.6; DB 22; Length 2077;  
Best Local Similarity 74.8%; Pred. No. 3e-56;  
Matches 416; Conservative 0; Mismatches 139; Indels 1; Gaps 1;

```

OY 2412 GCCGGCTCTTCTTGAAGTTCACGCCCTCTGCATGATGGCGGCCCAAGAGGAGATG 2471
DB 62 GCGGCTCTTCTTGAATTTTCCACCTGCTGAATTACTACGCCCCCAAGAGGAGATG 120
OY 2472 ATCGCCCTTCTTCTGAGTGTGGTGGCCATGGCGCTTACGACAAGAGG 2531
DB 121 ACCGCGCTTCTTCTGAGTGTGGTGGCCATGGCGCTTACGACAAGAGG 180
OY 2532 ACATCTCGGATTTCTCGAGTCCAAACCTGTGATGATGATGCAAGAGGATG 2591
DB 181 ACATCTCGGATTTCTCGAGTGTGGTGGCCATGGCGCTTACGACAAGAGG 240
OY 2592 CACACAGGCGCGCTCTTCTGAGTGTGGTGGCCATGGCGCTTACGACAAG 2651
DB 241 CTCACAGGCGCGCTCTTCTGAGTGTGGTGGCCATGGCGCTTACGACAAG 300
OY 2652 CTGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2711
DB 301 CAAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 360
OY 2712 GCAAAGTGAAGACCAATTAATCAAGTCAATTAATCAAGTCAATTAATCA 2771
DB 361 AGAAGTGAAGACCAATTAATCAAGTCAATTAATCAAGTCAATTAATCA 420
OY 2772 TTCTGTCTTCAATGATGATGATGATGATGATGATGATGATGATGATG 2831
DB 421 TCCTGTCTTCAATGATGATGATGATGATGATGATGATGATGATGATG 480
OY 2832 TTGGTTTCCCACTCACTATCACTCACTCACTCACTCACTCACTCACTCA 2891
DB 481 TTGGTTTCCCACTCACTATCACTCACTCACTCACTCACTCACTCACTCA 540
OY 2892 TGCTGGGCGGATCATGAGGCGGATCATGAGGCGGATCATGAGGCGGAT 2951
DB 541 TGCTGGGCGGATCATGAGGCGGATCATGAGGCGGATCATGAGGCGGAT 600
OY 2952 ATTTTGCCTGTGTGA 2967
DB 601 ACTTTGCATGTGATA 616

```

RESULT 12  
AAT21884  
ID AAT21884 standard; cDNA to mRNA; 301 BP.

AC AAT21884;  
XX  
DT 14-AUG-1996 (first entry)

XX Human gene signature HUMGS03426.  
DE  
XX Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KM cell typing; abnormal cell function; ss.  
XX

OS Homo sapiens.  
XX  
XX WO9514772-A1.  
XX  
XX 01-JUN-1995.  
XX  
XX 11-NOV-1994; 94WO-JP01916.  
XX  
XX 12-NOV-1993; 93JP-0355504.  
XX  
XX (MATS/) MATSUBARA K.  
XX (OKUBO/) OKUBO K.  
XX  
XX Matsubara K, Okubo K;  
XX  
XX WPI; 1995-206931/27.  
XX  
XX

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.  
PT for diagnosis of abnormal cell function, by preparing cDNA that  
PT reflects relative abundance of corresp. mRNA in specific human  
PT tissues  
XX  
XX Claim 1; Page 991-992; 2245pd; Japanese.  
XX  
XX

CC A single-stranded DNA (or its complementary strand or the corresp.  
CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
CC given in AAT19001-T76837 and which is able to hybridise to part of  
CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
CC sequences were obtained from 3'-directed cDNA libraries prepared  
CC from various human tissues; synthesis of cDNA was initiated from the  
CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
CC untranslated sequence is unique to a particular mRNA species, almost  
CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
CC is constructed so as to reflect accurately the relative abundance of  
CC different mRNAs in the particular tissue from which it was derived.  
CC The appearance frequency of a given GS in a cDNA library can be  
CC determined (esp. using primers and probes derived from the GS  
CC sequences) as a means of diagnosing abnormal cell function or for  
CC recognising different cell types.  
XX  
XX Sequence 301 BP; 96 A; 65 C; 63 G; 66 T; 11 other;

Query Match 6.6%; Score 285.2; DB 16; Length 301;  
Best Local Similarity 95.3%; Pred. No. 1.5e-48;  
Matches 287; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

```

OY 3519 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3578
DB 1 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 60
OY 3579 CTCACAGGCGCGCTCTTCTGAGTGTGGTGGCCATGGCGCTTACGAT 3638
DB 61 CTCACAGGCGCGCTCTTCTGAGTGTGGTGGCCATGGCGCTTACGAT 120
OY 3639 TAGCAGCAGGAGACGACACACACACACACACACACACACACACAC 3698
DB 121 TAGCAGCAGGAGACGACACACACACACACACACACACACACACAC 180
OY 3699 CCACACAGGAAACCTTGAAGAAATCACTTTCTAGAACGCGCTTACCT 3758
DB 181 CCACACAGGAAACCTTGAAGAAATCACTTTCTAGAACGCGCTTACCT 240
OY 3759 GTTATATATATATATATATATATATATATATATATATATATATAT 3818
DB 241 GTTATATATATATATATATATATATATATATATATATATATATAT 300

```

```
QY 3819 A 3819'
Db 301 A 301

RESULT 13
ID ABN97389 standard; DNA; 273 BP.
XX ABN97389;
AC ABN97389;
XX 13-AUG-2002 (first entry)
DE Gene #3887 used to diagnose liver cancer.
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumour; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS Homo sapiens.
XX WO200229103-A2.
XX 11-APR-2002.
XX 02-OCT-2001; 2001WO-US30589.
XX 02-OCT-2000; 2000US-237054P.
XX (GENE-) GENE LOGIC INC.
XX
XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX WPI; 2002-426119/45.
XX
XX Diagnosing and detecting the progression of liver cancer.
XX hepatocellular carcinoma or metastatic liver tumor in a patient,
XX involves detecting the level of expression of two or more genes in a
XX liver tissue sample.
XX
XX Claim 1; SEQ ID NO 3887; 298pp; English.
XX
XX The invention relates to a novel method for diagnosing and detecting the
XX progression of liver cancer, hepatocellular carcinoma or metastatic liver
XX tumor in a patient, and differentiating metastatic liver cancer from
XX hepatocellular carcinoma in a patient, involving detecting the level of
XX expression of two or more genes represented in ABN93503-ABN97455 in a
XX tissue sample. The method of the invention has hepatotropic, and
XX cytosstatic activity. The method is useful for diagnosing and detecting
XX the progression of liver cancer, hepatocellular carcinoma and metastatic
XX liver carcinoma in a patient. The method is useful for identifying
XX expression profiles which serve as useful diagnostic markers as well as
XX markers that can be used to monitor disease states, disease progression,
XX drug toxicity, drug efficacy and drug metabolism.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 273 BP; 59 A; 74 C; 62 G; 78 T; 0 other;

Query Match 6.4%; Score 273; DB 24; Length 273;
Best Local Similarity 100.0%; Pred. No. 4.5e-46;
Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4002 CAGGTTCTCCCTGCCAAAAGGGGGCTAGATGAGTGGTGGGCGCCGTGGACAGCTGAGA 4061
Db 273 CAGGTTCTCCCTGCCAAAAGGGGGCTAGATGAGTGGTGGGCGCCGTGGACAGCTGAGA 214
QY 4062 GTGGGATTTCATCAGACTCATGCAATTAACCTTTGATTGTTTCTTAAAGGAGACTCCCT 4121
Db 213 GTGGGATTTCATCAGACTCATGCAATTAACCTTTGATTGTTTCTTAAAGGAGACTCCCT 154
QY 4122 CGGCAAGATGGCAGAGGGTACGGAGTCTTCAGGGCCCGAGTTTCTACATTTAGCCATTGCA 4181
```

```
Db 153 CGGCAGATGGCAGAGGGTACGGAGTCTTCAGGCCAGTTTCTCATTAGCCATTGCA 94
QY 4182 GGGCTCCTTGTGGTGGATCAGAACTAATCCAGAGTGTGGGAAAGTGACAGTCAAAACCC 4241
Db 93 GGGCTCCTTGTGGTGGATCAGAACTAATCCAGAGTGTGGGAAAGTGACAGTCAAAACCC 34
QY 4242 CACCTGGAGCAAAATAAAAAACATACAAAACGT 4274
Db 33 CACCTGGAGCAAAATAAAAAACATACAAAACGT 1

RESULT 14
ABL64228/c
ID ABL64228 standard; DNA; 273 BP.
XX ABL64228;
XX 15-MAY-2002 (first entry)
DE Stomach cancer related gene sequence SEQ ID NO:2565.
XX
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
XX stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
XX cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
XX gene; ds.
XX
XX Homo sapiens.
XX WO200194629-A2.
XX 13-DEC-2001.
XX 30-MAY-2001; 2001WO-US10838.
XX
XX 05-JUN-2000; 2000US-209473P.
XX 05-JUN-2000; 2000US-209531P.
XX 18-SEP-2000; 2000US-233133P.
XX 18-SEP-2000; 2000US-233617P.
XX 20-SEP-2000; 2000US-234009P.
XX 20-SEP-2000; 2000US-234034P.
XX 20-SEP-2000; 2000US-234052P.
XX 22-SEP-2000; 2000US-234509P.
XX 22-SEP-2000; 2000US-234567P.
XX 25-SEP-2000; 2000US-234923P.
XX 25-SEP-2000; 2000US-234924P.
XX 25-SEP-2000; 2000US-235077P.
XX 25-SEP-2000; 2000US-235082P.
XX 25-SEP-2000; 2000US-235134P.
XX 26-SEP-2000; 2000US-235280P.
XX 26-SEP-2000; 2000US-235637P.
XX 26-SEP-2000; 2000US-235638P.
XX 27-SEP-2000; 2000US-235711P.
XX 27-SEP-2000; 2000US-235720P.
XX 27-SEP-2000; 2000US-235840P.
XX 27-SEP-2000; 2000US-235863P.
XX 28-SEP-2000; 2000US-236028P.
XX 28-SEP-2000; 2000US-236032P.
XX 28-SEP-2000; 2000US-236033P.
XX 28-SEP-2000; 2000US-236034P.
XX 28-SEP-2000; 2000US-236109P.
XX 28-SEP-2000; 2000US-236111P.
XX 29-SEP-2000; 2000US-236842P.
XX 29-SEP-2000; 2000US-236891P.
XX 02-OCT-2000; 2000US-237172P.
XX 02-OCT-2000; 2000US-237173P.
XX 02-OCT-2000; 2000US-237278P.
XX 02-OCT-2000; 2000US-237294P.
XX 02-OCT-2000; 2000US-237295P.
XX 02-OCT-2000; 2000US-237316P.
XX 03-OCT-2000; 2000US-237425P.
XX 03-OCT-2000; 2000US-237598P.
XX 03-OCT-2000; 2000US-237604P.
```



PR 03-OCT-2000; 2000US-237606P.  
 PR 03-OCT-2000; 2000US-237608P.  
 PR 01-NOV-2000; 2000US-244867P.  
 PR 01-NOV-2000; 2000US-245084P.  
 XX  
 XX (AVAL-) AVALON PHARM.  
 XX  
 PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 PI Soppet DR, Weaver Z;  
 XX  
 XX WPI; 2002-188264/24.  
 XX  
 PT Screening for anti-neoplastic agent involves exposing cells to a  
 PT chemical agent to be tested for anti-neoplastic activity, and  
 PT determining a change in expression of a gene of a signature gene set -  
 XX  
 PS Claim 1; SEQ ID 2565; 44p; English.  
 CC The present invention describes a method (M1) for screening for an  
 CC anti-neoplastic agent. The method involves exposing cells to a chemical  
 CC agent to be tested for anti-neoplastic activity, determining a change in  
 CC expression of at least one gene (I) of a signature gene set, where (I)  
 CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
 CC to ABL70110), or is at least 95% identical to (S), where a change in  
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
 CC activity and can be used in gene therapy. M1 can be used for screening  
 CC an anti-neoplastic agent, and can be used for producing a product which  
 CC is the data collected with respect to the anti-neoplastic agent as a  
 CC structure and/or properties of the agent. M1 can be used in the  
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
 CC carcinoma, papillary carcinoma and Wilm's tumour.  
 XX  
 XX Sequence 273 BP; 59 A; 74 C; 62 G; 78 T; 0 other;  
 SQ  
 Query Match 6.4%; Score 273; DB 24; Length 273;  
 Best Local Similarity 100.0%; Freq. No. 4.5e-46;  
 Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 4002 CAGGTTCTCCCTGCCAAAAAGGGGGCTAGATGAGTGGGCGCCCTGGACAGCTGAGA 4061  
 DB 273 CAGGTTCTCCCTGCCAAAAAGGGGGCTAGATGAGTGGGCGCCCTGGACAGCTGAGA 214  
 C 4062 GTGGGATTCATCCACATCACTATGCAATACCCCTTGTGTTGTTTCTAAAGAGACTCCCT 4121  
 DB 213 GTGGGATTCATCCACATCACTATGCAATACCCCTTGTGTTGTTTCTAAAGAGACTCCCT 154  
 QY 4122 CGGCAAGATGCGAGGGGTACGGAGTCTTCAGGCCAGTTTCTCACTTAGCAATTGCA 4181  
 DB 153 CGGCAAGATGCGAGGGGTACGGAGTCTTCAGGCCAGTTTCTCACTTAGCAATTGCA 94  
 QY 4182 GGGCTCCTTGGTGGGATGAGAACTATCCAGAGTGGGAAAGTGCAGTCAAAACC 4241  
 DB 93 GGGCTCCTTGGTGGGATGAGAACTATCCAGAGTGGGAAAGTGCAGTCAAAACC 34  
 QY 4242 CACCTGAGCAATAATAAAACATACAAACGT 4274  
 DB 33 CACCTGAGCAATAATAAAACATACAAACGT 1  
 RESULT 15  
 ABA20255  
 ID ABA20255 standard; DNA; 546 BP.  
 XX  
 AC ABA20255;  
 XX  
 DT 23-JAN-2002 (first entry)  
 XX  
 DE Human nervous system related polynucleotide SEQ ID NO 12586.  
 XX

KW Human; nootropic; neuroprotective; cytoskeletal; dermatological; virucide;  
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;  
 KW antiparkinsonian; antischizoid; antianemic; antiarthritic; cancer;  
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
 KW antiallergic; antidiabetic; antileuc; anticonvulsant; antifungal;  
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200159063-A2.  
 XX  
 PD 16-AUG-2001.  
 XX  
 XX 17-JAN-2001; 2001WO-US01334.  
 PF  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225477.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226686.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 20-OCT-2000; 2000US-0242221.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.

PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251889.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI WPI; 2001-541565/60.  
DR  
XX  
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system  
PT cancers and metastases -  
XX  
XX Disclosure; SEQ ID NO 12586; 1701pp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (ABA11004-ABA21534) and proteins  
CC (AB14678-AB18001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 546 BP; 133 A; 137 C; 171 G; 105 T; 0 other;

Query Match 5.9%; Score 254.6; DB 22; Length 546;  
Best Local Similarity 98.2%; Pred. No. 3.3e-42;  
Matches 268; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
QY 406 GGTGGAAAGCGGTGACACGCGCAAGAGCCCTGCGGTGATCTCCAAGTCCCATCCATGCC 465  
DB 2 GGTGGAAAGCGGTGACACGCGCAAGAGCCCTGCGGTGATCTCCAAGTCCCATCCATGCC 61  
QY 466 CCAGGACTCAGGCGCTCAGAGCTATTACCAATGGGACTTGGAGAGCGGAGTGAGCC 525  
DB 62 CCAGGACTCAGGCGCTCAGAGCTATTACCAATGGGACTTGGAGAGCGGAGTGAGCC 121  
QY 526 CCAGCCAGAGAGGGAGCCCTGCTGGGGGCGAGAGGGCGGGGCGGCGCTCCAGAGAGGAGA 585  
DB 122 CCAGCCAGAGAGGGAG-CCTGCTGGGGGCGAGAGGGCGGGGCTCCAGAGAGGAGA 180  
QY 586 GGTGCACTGAGACCCCTGCTGAAGCCTCAAGACAGTGGAAATGCTGCTGCACCCC 645  
DB 181 GGTGCACTGAGACCCCTGCTGAAGCCTCAAGACAGTGGAAATGCTGCTGCACCCC 240  
QY 646 CAAGAGGCGCGAGAGAGCCCTGCGAGAGAGCGGG 678  
DB 241 CAAGAGGCGCGAGAGAGCCCTGCGAGAGAGCGGG 273

Search completed: November 22, 2002, 09:12:08  
Job time : 616.737 secs